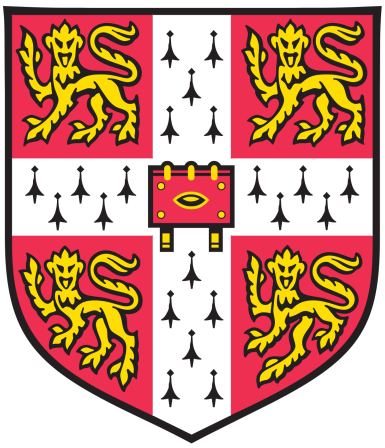
Techniques and Technologies for Earth-twin Discoveries

Richard D. Hall

Girton College

University of Cambridge



February 2020

A dissertation submitted for the degree of Doctor of Philosophy

Summary

In this Thesis I present and discuss the work completed during my three and a half years as a PhD student in the Exoplanet Research Group of the Cavendish Laboratory, University of Cambridge, UK. Most of my work has been in collaboration and partnership with the HARPS3 instrument and the Terra Hunting Experiment.

My focus is on the development of new techniques and technologies that are aimed at aiding the discovery of an 'Earth-twin' exoplanet. In the context of this work, I use the term Earth-twin to mean an Earth-massed rocky planet orbiting a Sun-like star at a period of around 300 days.

I created a pipeline prototype for fitting planetary models to radial velocity (RV) data. The data can contain any number of random or systematic noise sources, and can be poorly sampled. The analysis is conducted in a nested-sampling Bayesian framework and thus allows for the direct statistical comparison of different planetary models given some data set, and produces full posterior estimation for all the parameters of all the models. I used this analysis technique to test the feasibility of using intense ground-based RV surveys to detect Earth-twins, and to compare the results with typical survey cadences. I found that an intense survey reliably and regularly finds a variety of planets, including the Earth-twins, and out-performs the typical survey cadence.

The major new technology I have developed is an experiment to measure the geometric positions of the pixels of an optical CCD, and the data analysis pipeline to compute the results. In exoplanet science, precise measurements of the Doppler shift of the stellar spectral lines enable us to confirm the presence of planets. However, at some level of precision, our uncertainty of the *detector itself* starts to inhibit our detection capability. Hence, if we are to be successful in the discovery of low mass planets, we require knowledge of the sub-pixel structure of our detector. I used the analysis scripts to help plan and design an optical experiment which was then built to analyse a large format optical detector and measure the positions of the pixels. I found the simulation of the experiment can measure the pixel positions to a precision of less than 0.001 pixels, but the experiment was plagued with thermal variations and ultimately was not capable of such precise measurements.

Contents

List of Figures

List of Tables

Declaration

This work, in its entirety unless otherwise stated, is a product of my efforts during the past three and a half years as a PhD student in Experimental Astrophysics Group of the Cavendish Laboratory. Appropriate references, citations and thanks have been noted where external help or information was used.

This dissertation does not exceed the prescribed 60,000 word limit and 150 figure count limit for the Degree Committee of Physics and Chemistry.

Chapter 3 was published as 'On the Feasibility of Intense Radial Velocity Surveys for Earth-Twin Discoveries' in *Monthly Notices of the Royal Astronomical Society* (Hall et al. 2018).

Parts of Chapter 5 were published as 'Measuring the Effective Pixel Positions for the HARPS3 CCD' in *SPIE Proceedings High Energy, Optical, and Infrared Detectors for Astronomy VII* (Hall et al. 2016). The rest of Chapter 5 is, at the time of submission, part of an in–prep publication summarising the technique and the results.

Throughout the thesis, parts were submitted for my First Year report titled 'Measuring the Effective Pixel Positions of the HARPS-3 CCD', particularly from Chapters 1 and 4.

This document contains approximately 49,000 words excluding the appendix, captions, and references.

Acknowledgements

Thanks go, firstly, to my supervisor Didier Queloz for giving me this incredible PhD project. With his mentorship and guidance, my journey as a student has given me a vast variety of skills and experiences I would otherwise not have. I have had the opportunity to present my research across 3 continents to the academic elite, and countless times to the deserving public.

I thank my co-supervisor and friend Samantha Thompson. Her positive outlook and steely ability to weather the storm never ceases to amaze me, and it is safe to say that if any success comes out of the Terra Hunting Experiment it will largely be down to her. I would also like to thank the Exoplanet group at Cambridge for providing such a friendly and relaxed working environment and I wish good luck to all current and future students studying here.

I would also like to mention Joshua Briegal, Will Handley, and Vedad Hodžić for their endless help in debugging my often 'interesting' programming solutions. It is without doubt their help and tuition has been instrumental in my ability to succeed on this project.

The HARPS3 and Terra Hunting Experiment consortium meetings have provided many interesting and thoughtful debates over the years. I wish them luck and will be keeping a keen eye on the projects inevitable success.

Thanks to Greg Willatt and Dave Green for their support and help with running my code on the Astrophysics Cluster and for the LaTeX thesis template and help respectively.

For this research I have extensively used Python (Rossum 1995), and the open source packages of NUMPY, MATPLOTLIB, SCIPY, PANDAS, SEABORN.

Lastly I thank my close friends and family, and especially my beautiful fiancée, who has supported me effortlessly over the past 4 years. It is now time for me to return the favour.

Chapter 1

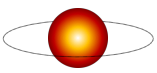
A Brief History of Exoplanets

> 'Somewhere, something incredible
> is waiting to be known.'
>
> *Carl Sagan*

It is quite rare that during one's lifetime an entire new *class* of astrophysical object is revealed to exist, as it happens on average once per century. In this chapter I will outline the young yet flourishing field of Exoplanetary Science, and illustrate where my research sits on this landscape.

1.1 Introduction to Exoplanets

The discovery of the first exoplanet orbiting a main sequence star, 51 Pegasi b Mayor & Queloz (1995), has ushered in a new era of astrophysics that has taken the community by storm. With evidence that the Solar System was not alone, astronomers have since been searching the sky to fully understand where it is placed in the cosmos. It is no coincidence that it has only been 24 years since the first exoplanet discovery: measuring the signal from a small, cool, faint rock is not easy. In this chapter I will outline the techniques that astronomers use to find exoplanets and give a brief history of the radial velocity technique. Due to the nature of my research, I place a heavy focus on this technique but mention others for completeness. A full description of the various techniques discussed, and others, can be found in Perryman (2014).



1.2 The Radial Velocity Technique

The radial velocity (RV) technique of exoplanet discovery is conducted by taking periodic spectra of a target star and tracking any Doppler shift present in the motion of the lines on a detector. Should a regular cyclical motion be found, it must mean that the target is moving regularly towards and away from the observer. In the absence of any obvious binary companions, the logical conclusion is that a planet is present. The RV technique is an indirect discovery method - an observer never sees the actual planet, only its gravitational influence on the star.

The RV technique was used to make the famous discovery of 51 Pegasi b in 1995, and has since been successful in discovering almost 20% of the nearly 4,000 known exoplanets (NExSci 2019). With the RV technique we measure a wavelength shift of the lines which corresponds to a source velocity, v_r, along the line of sight of an observer,

$$v_r = c\frac{\Delta\lambda}{\lambda} \tag{1.1}$$

where $\Delta\lambda$ is the measured wavelength shift of the line at wavelength λ, and c is the speed of light in a vacuum. We use this measured velocity and a model Keplerian system to estimate the parameters of the planets present in the system. Referring to Figure 1.1, the distance r from the barycentre of the system varies with,

$$r(t) = \frac{a(1-e^2)}{1-e\cos(\theta)}, \tag{1.2}$$

where a is the semi-major axis, e is the orbital eccentricity, and θ is the true anomaly. We write the true anomaly as a function of the eccentric anomaly given by

$$\theta = \frac{\cos E(t) - e}{1 - e\cos E(t)}. \tag{1.3}$$

The eccentric anomaly varies with time as,

$$E(t) - e\sin E(t) = \frac{2\pi}{P}(t - T_p) \tag{1.4}$$

where P is the orbital period and T_p is the time the planet passes through the periastron. Here we invoke Kepler's Law of Periods to relate the period to the semi-major axis with

$$P^2 = \frac{4\pi^2}{GM}a^3 \tag{1.5}$$

where G is the gravitational constant. As we are working in a relative frame of reference, M is the combination of the stellar ($\star$) and planetary (p) masses:

$$M = \frac{M_p^3}{(M_\star + M_p)^2}. \tag{1.6}$$

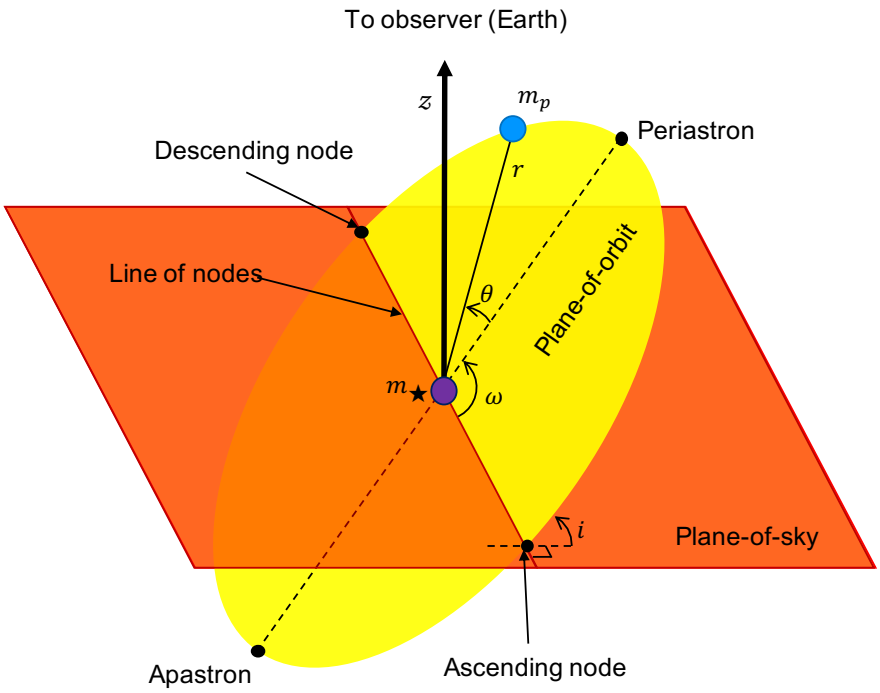


Figure 1.1: The 3D Keplerian system of a single planet with mass m_p, on a generic elliptical orbit around a star with mass $m_\star$. Where θ is the true anomaly, ω is the argument of periastron, i is the system inclination, and r is the distance between the planet and the star.

Looking at Figure 1.1, regardless of the inclination we only measure a velocity along the z axis towards the observer. Hence, we take the projection of motion from Equation 1.2 along the z axis:

$$z(t) = r(t) \sin i \cos(\omega + \theta). \tag{1.7}$$

Combining this, the RV of the star due to a planet along the line-of-sight of the observer is then:

$$v_r = z(t) = K[\cos(\omega + \theta) + e \cos \omega] \tag{1.8}$$

where K is the 'Radial Velocity Semi-Amplitude' and combines the physical constants from Kepler's 3rd Law as

$$K = \frac{2\pi}{P} \frac{a_\star \sin i}{(1 - e^2)^{1/2}}. \tag{1.9}$$

Here, we insert our expressions of P and M from Equations 1.5 and 1.6 to give

$$K = \left(\frac{2\pi G}{P}\right)^{1/3} \frac{M_p \sin i}{(M_\star + M_p)^{2/3}} (1 - e^2)^{-3/2}. \tag{1.10}$$

To see the effect of various values of ω and e on the RV curve, see Figure 1.2. The $M_p \sin i$ term demonstrates how it is not physically possible to measure the true mass of the planet, we

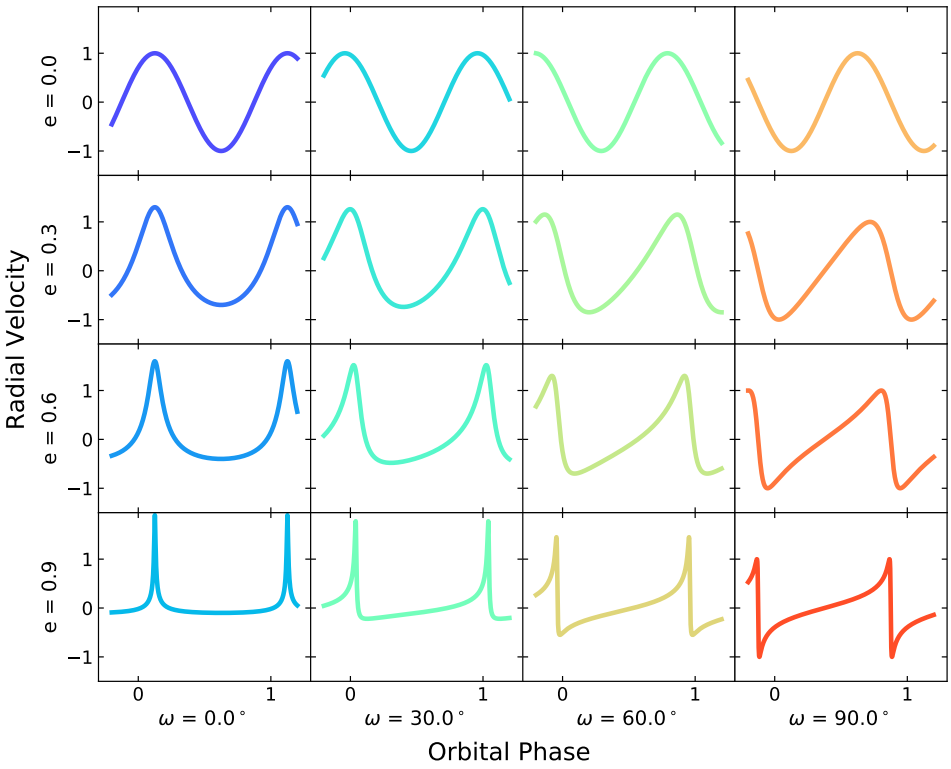


Figure 1.2: Examples of various RV curves from circular to highly eccentric (e) and with varying arguments of periastron (ω).

only obtain the minimum mass ratio of the two bodies projected onto the inclination plane. At this stage it is also worth noting that so far we have only considered a two-body system and any RV signal described here is purely sinusoidal. For a stable multi-planet system, such as our own Solar System, the total RV of the star due to planetary companions can be modelled as a sum of such signals with unique values for M_p, P, and e.

1.2.1 The Cross Correlation Function

The measurable from the RV technique is the motion of spectral lines on a detector. We convert the motion in units of pixels into a source velocity v_r from Equation 1.1. The conversion rate is unique to each spectrograph as it is a function of pixel size, the total number of pixels, spectral resolution, and the wavelength range imaged. For a modern high resolution spectrograph the resolution is typically of the order $R = 100,000\lambda/\Delta\lambda$, (Mayor et al. 2003), it observes a spectral range of $380 < \lambda < 690$nm, and it samples a calibration spectral feature FWHM (full-width half-maximum) with around three pixels. With this optical configuration, a measured

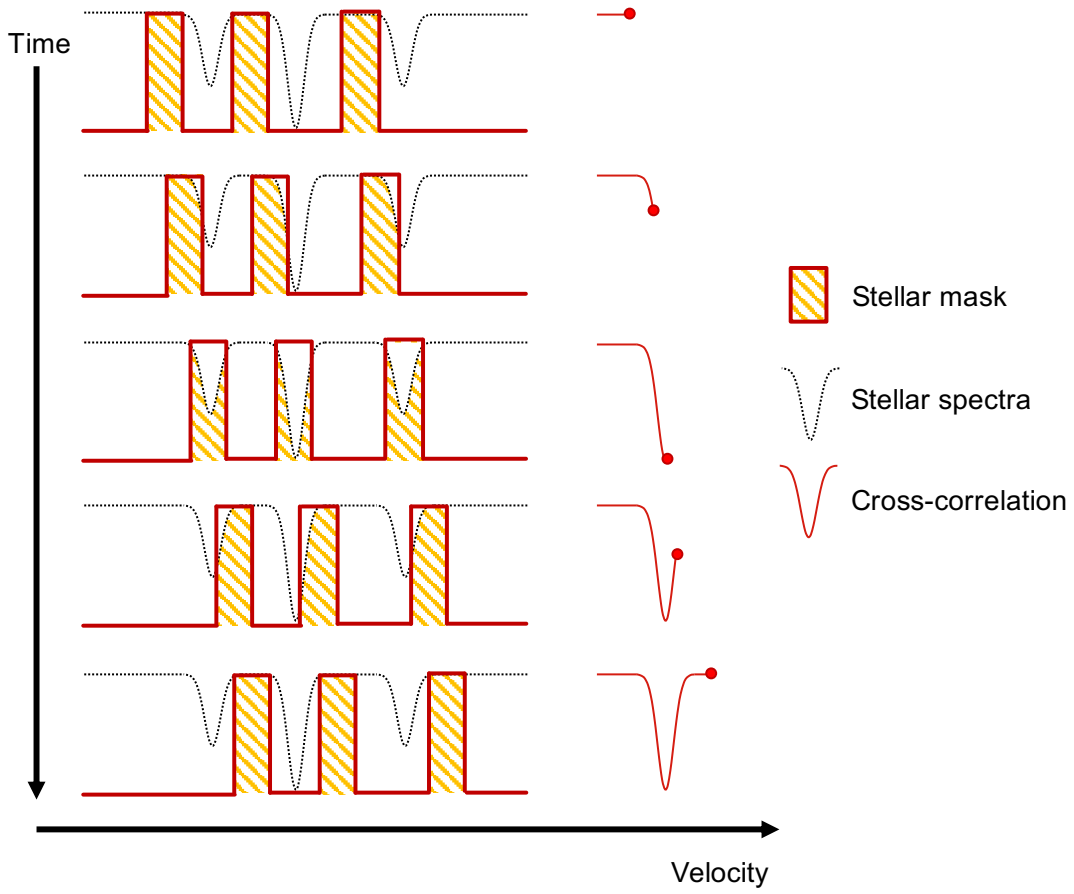


Figure 1.3: A cartoon of how to build a cross-correlation function from a stellar spectra and a stellar mask. The mask is stepped across the spectra and the product is taken resulting in a smooth Gaussian in the ideal case. The velocity at the peak of the Gaussian is the measured stellar velocity at that time and the process is repeated for future observations.

shift of one FWHM corresponds to $\sim$3000 m s^{-1} of source velocity, whilst a 1 m s^{-1} velocity measurement means the spectrum has moved only 0.001 pixels across the detector. The stellar spectra will contain many thousands of spectral lines for the observer to track. We can average the motion across all of these to help beat down white noise, and to increase our measurement precision on the spectra as a whole.

To obtain the actual velocity measurement from the stellar spectra, a cross-correlation function (CCF) between the measured spectra and a template is taken. This reduces the entire stellar spectra to a curve that can be fitted by a single Gaussian with a peak velocity. A different template spectra is used for different spectral types to ensure the template best matches that of the real spectra. As seen in Figure 1.3, the mask is stepped across the spectra and the product is taken. For lines of different widths and depths, the mask can be modulated to yield better precision on the CCF fit (Pepe et al. 2002). For each new spectrum taken on subsequent

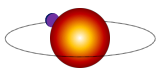


Table 1.1: A sense of scale when discussing RVs is useful. This table shows the RV of a Solar mass star from various planets at various orbital distances for $e = 0$ and $i = 90°$.

Planet	Orbital Separation (AU)	Period (days)	K_1 ($m\,s^{-1}$)
Jupiter (317 $M_\oplus$)	0.1	12	89.8
	1.0	365	28.4
	5.0	4081	12.7
Neptune (17.2 $M_\oplus$)	0.1	12	4.8
	1.0	365	1.5
Super-Earth (5 $M_\oplus$)	0.1	12	1.4
	1.0	365	0.45
Earth (1 $M_\oplus$)	0.1	12	0.28
	1.0	365	0.09

observations, the stellar velocity will have shifted by some small amount should there be an underlying Keplerian signal. The CCF peak position as a function of time is what is plotted for RV curves as seen in Figure 1.2.

1.2.2 The History of Radial Velocity Methods

With a description of the physics of a planetary system in place, we can begin to look at various key instruments that have operated over the past few decades and discuss the technological milestones that have driven the RV measurement precision down to the $\sim$10 $cm\,s^{-1}$ level. For reference, Table 1.1 puts this RV value into perspective with other planets orbiting a solar mass star.

Astronomical spectroscopy has been a staple tool for exoplanet hunters over the past three decades, but has been used for many more decades prior by other astronomical disciplines such as galaxy dynamics, stellar chemistry, and nebulae compositions. Notable RV instruments include ELODIE (Baranne et al. 1996), and its sister instrument CORALIE, which have a higher resolution detector and improved temperature stabilisation (Queloz et al. 2000). HIRES (Vogt 1992) installed on the Keck utilised a high-resolution CCD detector capable of imaging in a broadband range of 300 nm to 1100 nm. Later, HARPS (High Accuracy Radial velocity Planet Searcher), was designed to bring RV precision up to a new order of magnitude of 1 $m\,s^{-1}$ (Mayor et al. 2003) and was followed up with HARPSN (Cosentino et al. 2012) to conduct a similar survey in the Northern hemisphere. Another high-precision instrument is the Automated Planet Finder (APF, Vogt et al. (2014)) which is a high-cadence automated search for exoplanets orbiting nearby stars. Jumping another order of magnitude in precision to 10 $cm\,s^{-1}$, ESPRESSO (the Echelle Spectrograph for Rocky Exoplanets and Stable Spectroscopic Observations, Pepe et al. 2014) has been installed on the VLT (Very Large Telescope) and is operational as of

October 2018.

As mentioned in Section 1.2.1, the principle of Doppler spectroscopy relies on capturing precise spectra from the target star over a long time period. Should the star be moving relative to an observer, a Doppler shift in the spectral lines will be measured. There are a number of reasons why the star may appear to be moving, such as various types of stellar activity discussed later in Section 3.2, but a common source of Doppler shift is due to the cyclical motion of co-orbiting bodies.

Doppler spectroscopy was initially used to investigate spectroscopic binary stars - a binary system where the two bodies are sufficiently close that they cannot be resolved with a traditional imaging telescope (Hilditch 2001). With such a companion, the gravitational interaction between the two bodies causes them to orbit a common centre of mass and thus will move towards and away from an observer during the course of one orbit. For a binary star system, the magnitude of this RV is of the order $10\,\mathrm{km\,s^{-1}}$.

Should the star have a planetary companion however, they will both still orbit a common centre of mass but due to the vast mass disparity between the bodies this point is usually close to the centre of the star. In this scenario, the RV due to a large planet may only be around $1\,\mathrm{m\,s^{-1}}$ to $100\,\mathrm{m\,s^{-1}}$. The first feasible proposal to find exoplanets came from Otto Struve who suggested that we may just detect a Jupiter–sized planet at 0.02 AU with 'the most powerful Coudé Spectrographs in existence' (Struve 1952). For the aforementioned discovery of 51 Pegasi b in 1995, the technology of the then state-of-the-art instrument ELODIE allowed the astronomers to make the precise measurements of the star to discover the planet with an RV semi-amplitude of $59 \pm 3\,\mathrm{m\,s^{-1}}$ over a period of just 4.2 days. Whilst this might seem impressively precise, modern RV instruments are now pushing instrument precision to the $10\,\mathrm{cm\,s^{-1}}$ limit. This goal is common among competing instruments as it is the RV semi-amplitude for an Earth-mass planet orbiting a Solar-mass star at 1AU.

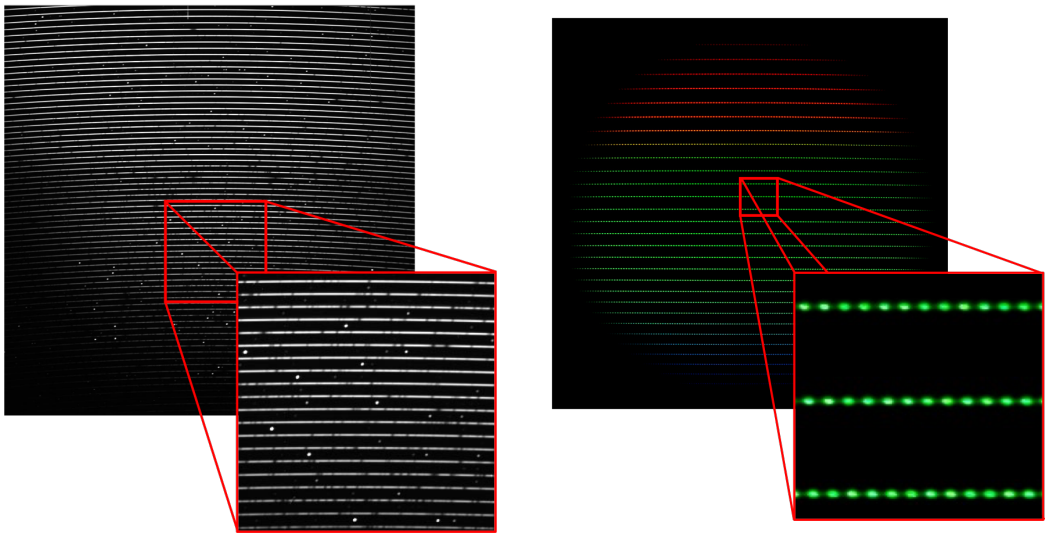
At the turn of the millennium, ELODIE offered many new technological improvements over traditional spectrographs (Baranne et al. 1996). Along with AFOE (Advanced Fiber Optic Echelle, Brown et al. 1994), ELODIE demonstrated that precise control of the entire optical system was crucial in the search for low RV amplitude signals. Two major new technologies employed were optical fibers to stabilise the stellar spectra on the detector (Queloz et al. 1999), and a simultaneous reference spectrum. Both of these helped ELODIE reach a precision of $10\,\mathrm{m\,s^{-1}}$.

Optical fibers exhibit a useful passive process known as 'scrambling' - their output beam is nearly independent from their input. This has the effect of reducing the illumination variations caused by thermal vibrations and drift that the instrument may be subject to, and thus reducing the overall RV uncertainty. Optical fibers also partially decouple spatial motion between the

telescope and the spectrograph which means that any spectral drift is likely to originate at the source rather than in the instrument. Prior to fibers, slit-illuminated spectrographs only decoupled one of the two axes of motion between the detector and the telescope, which meant that instrumental drift (such as temperature or vibrations) could still influence the spectra and was a major limitation for early spectrographs.

Another technology that ELODIE employed was the use of a simultaneous reference spectra that would be imaged along with the target spectra. The reference, or calibration, spectrum allows for instrumental drift and any local variations to be measured when obtaining a velocity of the target star. If you have a known spectrum simultaneously projected onto the detector with a target spectrum, you can track the relative differences to correct for any local anomalies or drifts. An example of a simultaneous reference spectrum can be seen in Figure 1.4a. Previous instruments would use an absorption cell, usually containing iodine, placed in the beam line to superimpose known spectra lines on top of the stellar spectra. When the full set of images were analysed, the temporally stable and known lines could be subtracted along with any motion they may exhibit due to instrumental drift. This technique, however, would result in flux losses of up to 15%, and the cells would slowly degrade with age. A detailed model of this spectra is also needed to disentangle the two from each other. Also, the iodine absorption lines do not cover the full spectral range as required by optical spectrographs. ELODIE's simultaneous spectra came from a Thorium-Argon (ThAr) lamp whose light was passed down a parallel fiber into the dispersing optics of the instrument and eventually onto the CCD. Typically a user may just take one calibration spectra at the start and end of a night, but it is possible to have one in every exposure. ThAr lamps are a major improvement over Iodine cells as the lines have an *absolute* position in wavelength space as they are determined by the atomic properties of the gas. However, ThAr catalogues are required for each instrument as each lamp will have unique impurities and degradation which creates a unique spectra.

Around the same time, a second technique for simultaneous reference was developed - a Fabry-Perot interferometer for a spectrograph (McMillan et al. 1993). A Fabry-Perot interferometer is created from two partially reflective surfaces facing each other with light entering this cavity at an angle. As the light bounces up this cavity, another beam can escape out of one side only. At each successive reflection a series of parallel beams are created each with a path difference proportional to the angle of incidence and the amount of separation between the reflective surfaces. The parallel beams can be focused to a point and the intensity will be the result of constructive or destructive interference. In a spectrograph the Fabry-Perot creates very sharp intensity peaks at a predetermined pitch in wavelength space, an etalon, which effectively acts as a ruler to measure the drift of the stellar spectra, see Figure 1.4b. Whilst these calibration sources are ideal when approaching 1 m s^{-1}, they require intense illumination as the



(a) ELODIE with ThAr calibration

(b) HARPSN Fabry-perot etalon

Figure 1.4: An example of the differences between the ThAr and Fabry-Perot calibration sources. The irregularly spaced and uneven lines from the ThAr lamp (interleaved with the target spectrum) are vastly inferior to the regular etalon from the Fabry-Perot.

light is spread across many thousands of peaks. This thermal energy can also be the cause of unpredictable thermal drift of the etalon itself (Wildi et al. 2011). To use a Fabry-Perot for sub-1 m s^{-1} precision measurements, a lot of work has to be conducted to stabilise the cavity and frequency-lock the spectra to prevent unpredictable drifts. Recent developments have even claimed ~1 mm s^{-1} peak tracking on the reference spectra (Reiners et al. 2014).

One final reference spectrum technology to discuss is the Laser Frequency Comb (LFC) (Murphy et al. 2007). An LFC produces a similar etalon to the Fabry-Perot but is more precise as it is reference from atomic clocks, plus it offers greater lifetime due to the nature of its construction being largely from off-the-shelf optical components.

The laser produces a small pulse that is sent on a repeating optical path. At each repetition another pulse can be emitted which eventually results in a long train of optical pulses at specific timings. In an ideal Fourier space, these pulses make up the comb as a series of delta functions with a specific frequency spacing. In reality, the comb features are broadened and it has a global envelope which limits the minimum and maximum wavelength at which these peaks can be emitted. Early precision estimations of an LFC place line tracking errors at the order of 5 cm s^{-1} (Li et al. 2008), and in practise HARPSN has reported short term repeatability to 2.5 cm s^{-1} (Messenger 2012).

To summarise the relative pros and cons of each of the discussed calibration sources I have updated the comparison table found in Murphy et al. (2007), see Table 1.2. For the LFC specifically, it is possible to obtain acceptable intensity across the spectral range with careful

Table 1.2: The requirements for an ideal wavelength calibration source updated from Murphy et al. (2007). 'Poss.' means it is possible depending on the design and construction of the specific calibration source and instrument used.

Calibration Requirements	ThAr	I_2 cell	Fabry-perot	LFC
From fundamental physics	Yes	Yes	No	Yes
Individually resolved	Mostly	Yes	Poss.	Yes
Uniformly spaced	No	No	Yes	Yes
Cover optical range	Yes	No	Poss.	Yes
Uniform intensity	No	No	Poss.	Poss.
Long-term stability	No	Poss.	No	Yes
Maintain Object S/N	Yes	No	Yes	Yes
Exchangeable	Yes	Yes	Poss.	Yes
Easy to use	Yes	Yes	Poss.	Poss.
Low Cost	Yes	Yes	Yes	Poss.

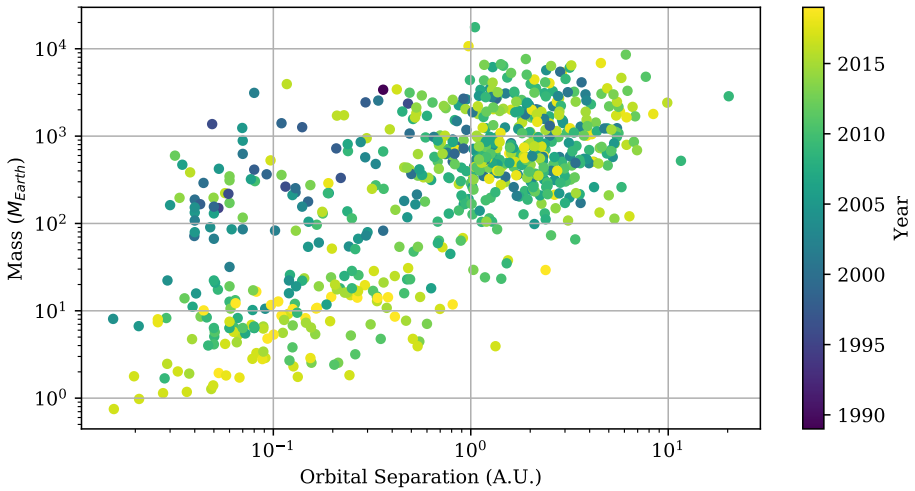


Figure 1.5: The mass and orbital separation of all exoplanets detected with the radial velocity technique, colour-coded as per their year of discovery - darker is an earlier discovery, brighter is more recent.

instrument design as seen by the implementation from ESPRESSO.

The progression of instrument precision and calibration has improved our detection capabilities as time progresses, see Figure 1.5. Here, there is a clear trend that lower mass and longer-period planets are being found more recently with the RV technique. It is also clear that there is a diagonal boundary from lower-left to upper-right under which no planets have been found with the RV technique. It is in this region where we would find an Earth-twin orbiting the sun at 1AU.

1.3 Other Detection Techniques

In this section I briefly discuss three other techniques for exoplanet discovery but there are others that I will not mention. The interested reader may refer to Perryman (2014) for more information.

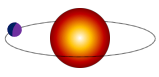
1.3.1 The Transit Technique

When the orbital plane of an exoplanetary system aligns with an observer's line of sight, a periodic dip in brightness is observed as the planet passes across the disc of the star and blocks some of the light, Figure 1.6. We call this event a transit, and it was first used to observe the exoplanet HD 209458b by Charbonneau et al. (2000).

As mentioned, the RV technique only provides an estimate of the lower bound of the planet's mass, and a good estimate on the period. However the transit technique provides a good estimate on the radius ratio between the planet and the star, and a good estimate of the period. Using a temperature-luminosity relationship and the Stefan-Boltzman law, the radius of the star can be well estimated to within a 10% which helps provide bounds on the radius of the transiting planet.

Combining these the two techniques allows for *both* mass and radius measurements to be made. The transit also greatly confines the system inclination to a narrow range which massively increases the accuracy of the mass estimate. From here, the bulk density and geological compositions can be inferred (Seager et al. 2007). Another useful implementation of the transit technique lies with transmission spectroscopy. Here, the transit can be observed in multiple narrow bandwidths and different transit depths may be observed. From this, the various atmospheric components may be detected as they will preferentially absorb different amounts of flux at different wavelengths. This leads to the potential for assessing the habitability of exoplanets by detecting key biomarkers in their atmospheres (Seager et al. 2016). However, the current state of these investigations seems to be limited by the presence of clouds and haze in the upper layers of the planet atmosphere that inhibit the transmission of these biomarkers (Sing et al. 2016), but recent results are providing evidence of water features after detailed spectra of WASP-39b were taken with the Hubble Space Telescope (Wakeford et al. 2018).

One potential downside to the transit technique stems from the very nature of the method itself. An observer only obtains useful planetary information during the few hours of the transit. For all other points on the planetary orbit, the measured signal is just the photometric variability of the star. You have to be lucky to find a long-period planet as unless you monitor the star for months on end, you may miss the transit. This is why observers have migrated to a wide-field paradigm for their surveys.



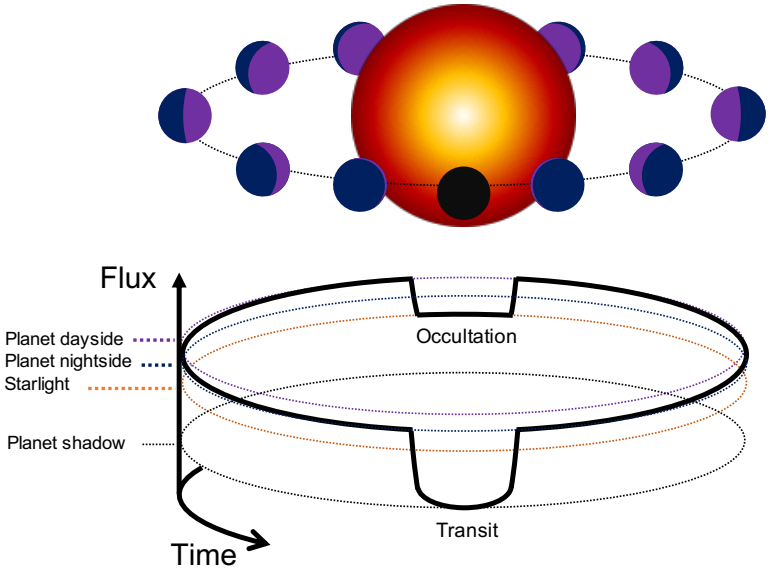


Figure 1.6: The light-curve of a single transit as a function of time. The subtle differences in the brightness of the day and night sides of the planet can be measured with enough photometric precision. The transit occurs when the planet passes in front of the star whilst the occultation is when the planet passes behind the star. This plot is inspired by Winn (2010).

Notable ground based wide-field surveys include HATNet (Bakos et al. 2002), WASP (Street et al. 2003), NGTS (Wheatley et al. 2018). Targeted transit surveys are also used such as with the SPECULOOS project (Burdanov et al. 2018). A wide-field survey generally comprises multiple wide-angle telescopes that image a large section of the sky and monitor many hundreds to thousands of stars simultaneously. The data pipelines are designed to track the photometry of each star and then analyse the light curve to detect any potential transits. A ground-based survey has the advantage of being relatively cheap (even amateur level optics at the right observatory can produce incredible results) and they can cover a large section of the sky to detect as many planets as possible. However, they are limited in their ability to accurately track the moving night sky and in the seeing conditions that can vary greatly throughout the year.

To overcome these issues the solution is to go to space. Despite coming at a great cost, space-based wide-field surveys have witnessed incredible success. The Kepler spacecraft has found the majority of known exoplanets to date and had nearly four years of continuous viewing of a fixed field in the sky (Koch et al. 2010). At the time of writing, of the 3,916 known exoplanets Kepler has discovered 2,337, and there are another 2,424 promising candidates to be vetted. The Kepler mission was transformed into K2 when the spacecraft lost its fullpointing ability due to 2 out of 4 reaction wheels failing (Howell et al. 2014). Since then, K2 has found 359 planets with another 472 candidates awaiting confirmation (NExSci 2019).

A more recent wide-field survey mission is TESS (Transiting Exoplanet Survey Satellite).

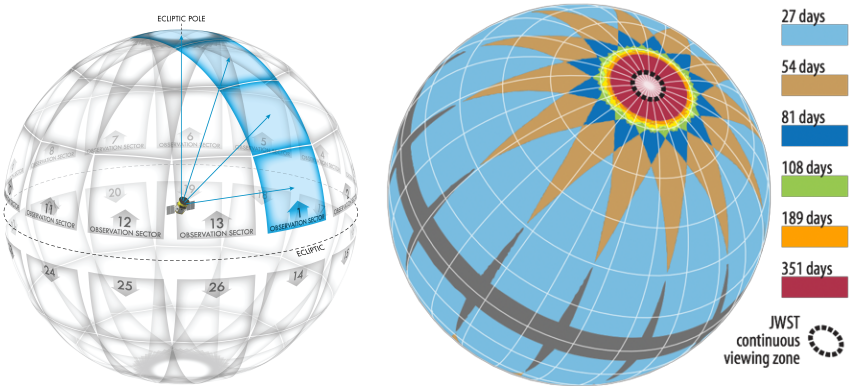


Figure 1.7: A schematic of the TESS observation pattern showing how the fields overlap, Figure from Ricker et al. (2014).

TESS is an all-sky survey and its primary two year mission will see it monitor the nearby stellar neighbourhood for any transiting signal. The TESS fields are only observed for 27 days, but many overlap to give a combined duration of up to 351 days, see Figure 1.7. TESS will then flip to the southern hemisphere and repeat this observation pattern and will collect data on over 200,000 stars and will likely detect many thousands of new exoplanets (Ricker et al. 2014).

1.3.2 The Direct Imaging Technique and Microlensing

The last two techniques to mention for completeness are direct imaging and gravitational microlensing. These more recent techniques have not yet returned discoveries in the same quantities as RVs and transits, but they make use of promising technologies that could take up the mantle in the near future.

Direct Imaging (DI), as the name suggests, is a technique where the planet and the star are spatially resolved on the detector and the planet is directly imaged. Efforts are made at the hardware and software level to mitigate the light from the host star which reveals the faint glow of the orbiting planets. At the time of writing, the closest planet detected with the DI technique is Beta Pic b which lies at 9 AU from the star (Lagrange et al. 2009). DI is a direct detection technique which means the photons collected have come from the planet itself. This gives the possibility for direct atmospheric and temperature measurements of a planet during its orbit.

Gravitational lensings are one-off chance events. The principle is the same as when a large body massively distorts some background galaxies, but on a much smaller scale. They occur when a planet hosting system passes in front of a distant star and the slight gravitational bending of the light gives an observer two small peaks of brightness: one for the main star and an even smaller peak for the planet. During these events it is possible to estimate the mass and orbital separation of the planet. This technique is sensitive to planets of just a few Earth-masses, but

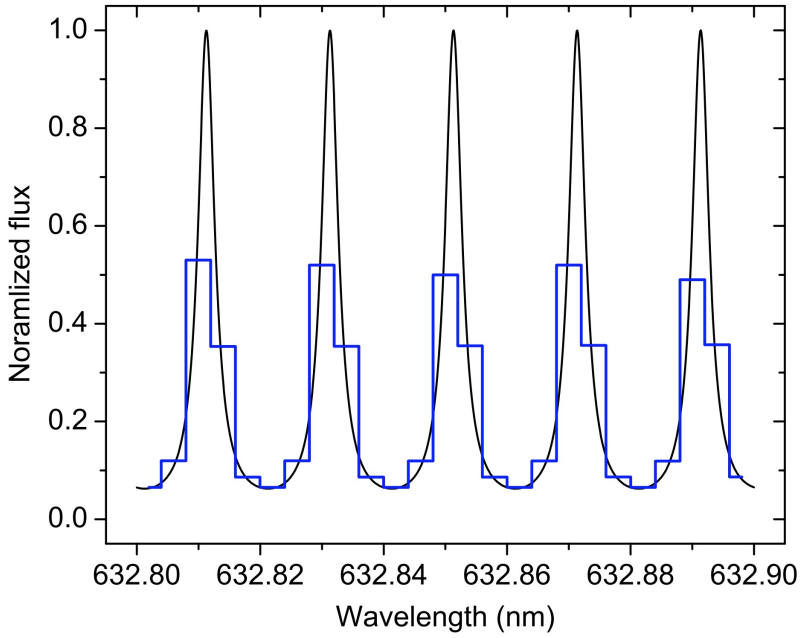


Figure 1.8: An example of how the resolution of a CCD affects the quality of calibration data collected. Here a Fabry-Perot etalon (black) is being imaged by a detector of discreet pixels (blue). Any pixel displacements or intensity variations between the pixels would yield an incorrect measurement of the true positions of the etalon, and hence the calibration would be inherently flawed, Figure from Reiners et al. (2014).

also at a few AU in separation (Beaulieu et al. 2006).

1.4 A Word on Detectors

As a large section of this research concerns the precise characterisation of optical detectors, I briefly introduce and discuss them here. I go into much more detail in Chapter 4.

CCD sensors (Charge-Coupled Device) are the preferred detector of choice over CMOS sensors (Complementary Metal Oxide Semiconductor) for astronomy. CCDs offer much higher sensitivity and uniformity due to how they are fabricated (Janesick 2001). CMOS sensors have an individual amplifier and readout circuit per pixel. Whilst this greatly increases readout speed it comes at a cost of photometric precision, gain non-uniformities between pixels, and that up to 75% of the imaging surface of a single pixel is used by this circuitry and thus does not contribute to collecting useful photons.

For high resolution spectrometry or photometry, the ideal scenario is to perfectly capture the spatial and temporal intensity variations that land on a detector and feed them into an analysis pipeline. This could result in either perfectly reconstructing the profile and motion of spectral lines for RV measurements, or the small fluctuations of light from a small transiting planet.

However, at some level, all imaging sensors exhibit defects or inhomogeneities which affect the quality of data being collected.

The pixels of a CCD are manufactured to very precise specifications, but they are still subject to random and systematic errors both in their physical position, size, and in their photometric response to light. Whilst measures to mitigate the photometric effects are commonplace in astronomy, the deviation of their position to a perfect grid will yield errors in a discipline which requires precise positional measurements. For RVs, these pixel displacements can affect how the science spectra and the calibration spectra are measured, e.g. (Cersullo et al. 2019), and will also introduce time-dependent errors as the spectra moves across the detector. To give a sense of scale, a 1×10^{-3} error of a pixel's position can return almost $3\,\mathrm{m\,s^{-1}}$ offset in the radial velocity per spectral line. Figure 1.8 shows how a Fabry-Perot etalon is imaged, and how one could easily imagine that a slight error of the pixel position or response could manifest in a large uncertainty of the calibration. The need for sub milli-pixel CCD calibration is necessary if we are to take advantage of the incredible instruments and calibration techniques now provided by the community, and to truly have the capability to discover Earth-twin planets.

1.5 Why Haven't we found any Earth-twin planets?

With the development of RV instruments reaching a stage where we can physically measure a $10\,\mathrm{cm\,s^{-1}}$ signal, you may wonder why Figure 1.5 does not have more samples around $1M_{\oplus}$ and 1AU. The issue is complex and multifaceted, and this thesis only aims to tackle a subset of them.

The first challenge to overcome is one of engineering. We need instruments capable of reliable $10\,\mathrm{cm\,s^{-1}}$ precision, installed on telescopes that give us enough photons for our target stars such that we obtain good quality spectral data. We also require precise calibration from these instruments and their detectors.

Secondly, we need to sample the stars such that we have the required density and duration of data. For an Earth-twin with a period of around 300 days, an observer ideally needs to schedule their observations evenly throughout the calendar year and would want to observe at least a few orbits in total. It is not guaranteed that this level of observation time is readily available on large telescopes.

Thirdly, the RV of a star is never a simple Keplerian sinusoid. The star exhibits many physical phenomena on different periods and amplitudes which manifest as multiple false positives during data analysis. I will discuss these in detail in Chapter 3.

Lastly, we need some luck. The RV gives an estimate of the mass ratio of the system projected onto the inclination relative to us, if the system is too inclined we would not measure

a signal at all. We also need to be fortunate enough that there even *is* an Earth-twin to be discovered in our sample of stars. Early indications from the Kepler sample suggest that Earth-mass planets are relatively common (Zeng et al. 2018), and that Sun-like stars likely have multiple planet systems with a 36 ± 14% chance of a planet orbiting in the habitable zone (Mulders et al. 2018). But for these stars the occurrence rate of an Earth-twin *in* the habitable zone may only be between 1-3% (Catanzarite & Shao 2011; Foreman-Mackey et al. 2014; Winn & Fabrycky 2015). If these estimates are correct, or even optimistic to within an order of magnitude, an Earth-twin hunting program would require a sample size of between 50-100 Sun-like stars to be confident for a discovery. Such a program, even with the best instrument available, would still need to collect many hundreds of RV measurements per target over a duration of many years if they are to be confident they have not missed the $10\,\mathrm{cm\,s^{-1}}$ signal, or that they have not confirmed false positives.

We have, however, found multiple planets *similar* to the Earth. The first terrestrial candidates were two superearths found to orbit CoRoT-7 (Léger et al. 2009; Queloz et al. 2009) at 0.85 and 3.69 days, clearly too close to their star to be habitable. Later, Kepler found a terrestrial candidate in the habitable zone of the M1 dwarf star Kepler-186. The fifth planet of this system (Kepler-186f) has a radius just 10% larger than the Earth and orbits Kepler-186 at 129.9 days but receives just 32% of the flux as compared to the Earth due to the star being a lot cooler than our Sun (Quintana et al. 2014). Lastly, the TRAPPIST-1 system was found to host *seven* Earth-sized planets with radii between 0.76 to 1.23 $R_\oplus$ in short orbits between 1.5 to 20 days of the host star, a small M8 dwarf (Gillon et al. 2017). Even with TRAPPIST-1 where many of the planets receive a similar irradiation to the Earth and have equilibrium temperatures suited for liquid water, they still may not be truly habitable. M dwarfs are known for their violent flares and outbursts that occur frequently and would likely damage any living organism on the planet should one strike the surface. Therefore, on the topic of habitability, we must look at Solar-type stars and try to find Earth-like planets on orbits of a few hundred days.

The next obvious question would be to ask 'Well, has there been a multi-year program taking precise RV measurements of 50 nearby Sun-like stars to find Earth-twins?...'

Chapter 2

HARPS3 and the Terra Hunting Experiment

'In the beginning the Universe was created. This has made a lot of people very angry and been widely regarded as a bad move.'

Douglas Adams

The work conducted in this thesis is in collaboration with the HARPS3 spectrograph and the Terra Hunting Experiment. The HARPS3 instrument is a new RV spectrograph scheduled for first-light in 2021, and the consortium that is designing and building it are also proposing a 10-year observational program called the Terra Hunting Experiment which will be the first dedicated RV survey designed specifically to hunt for Earth-twin Exoplanets. In this brief chapter I will give an overview of both of these projects and demonstrate that the specifics of my work are needed for both of these projects to succeed, and that this work is also relevant for the greater field of Exoplanets.

2.1 The Need for a New HARPS

As the name suggests, HARPS3 is the third and most recent in the series of the HARPS (High Accuracy Radial velocity Planet Searcher) spectrographs. A complete description of the

instrument and more information on the Terra Hunting Experiment can be found in Thompson et al. (2016) on which much of this chapter references.

With a design almost identical to HARPS and HARPSN, you may wonder why astronomers would choose to build a third when they could be more adventurous and push the technology further. The HARPS instruments have shown incredible stability and 1 m s^{-1} precision consistency over the past decade and their robust design has stood the test of time (Lovis et al. 2006b). If a new design were commissioned, there is no guarantee that it would survive the required duration of a program hunting for Earth-twins. In addition to this, the design and construction costs are minimised when using plans that already exist. This massively increases the rate of construction and allows a team to focus on intelligent use of the instrument along with improved software systems and calibration systems. It is very much a mantra of 'if it ain't broke, don't fix it'. The HARPS3 consortium are also planning to roboticise and automate the operations of the Isaac Newton Telescope, which will fully overhaul how this facility is operated.

2.1.1 The HARPS3 Instrument

HARPS3 will be a fiber-fed high resolution spectrograph with a resolution of $R \simeq 115,000$, a spectral range of $380 < \lambda < 690$nm, and will be installed on the 2.5 m Isaac Newton Telescope (INT) in La Palma in the Canary Islands, see Figure 2.1.

Two optical fibers will provide simultaneous measurements of the target spectrum and the reference spectrum. There is also a polarimeter which allows for an alternative yet complimentary method of monitoring stellar activity.

The instrument will be housed in a series of thermal chambers that have increasing temperature stability culminating with the third and innermost chamber reaching ±10 mK stability. Inside, the vacuum vessel houses the grating optics, lenses, camera system and the detector. A continuous-flow cyrostat will feed liquid nitrogen to the detector to maintain the temperature at 77 K which minimises thermal noise and drift. HARPS3 also requires a new Cassegrain adapter to be installed at the Cassegrain focus of the INT. This unit will contain the tip/tilt system for target tracking, the calibration source optics, the polarimeter, and the fiber selection mechanism.

The detector is the same model as used by HARPSN: a 4096×4096 e2V CCD[a] but with an additional thin film coating that boosts the quantum efficiency in the bluer end of the spectral range. As I will discuss in great detail in Chapter 5, a large portion of the calibration of HARPS relies on a precise pixel mapping of the detector. In Section 1.2 I discussed how a 1 m s^{-1} motion of a spectral line is just 0.001 pixels across the detector. If HARPS3 and the Terra Hunting Experiment is to be successful in confirming the presence of 0.1 m s^{-1} signals, we must

[a] CCD Specifications can be found here

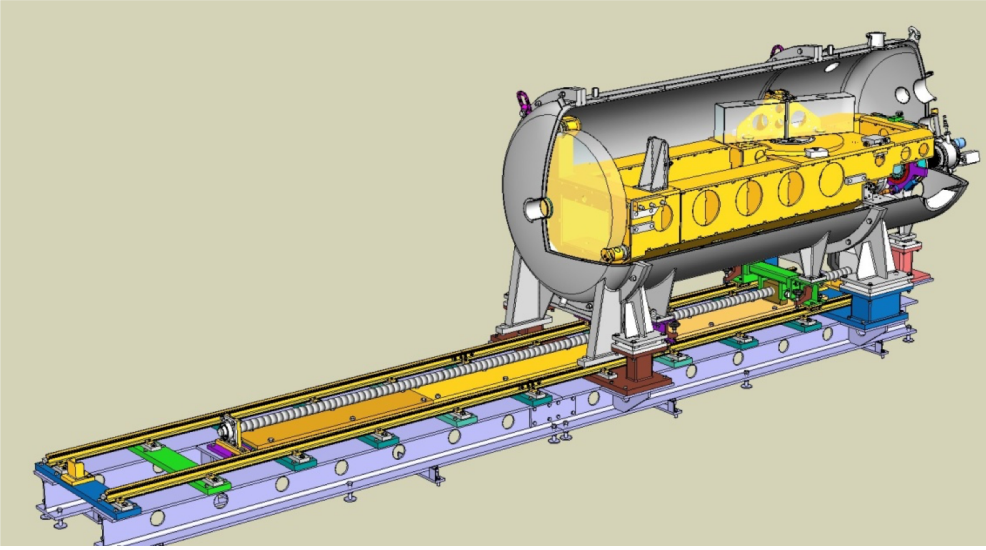


Figure 2.1: A render of the current status of the HARPS3 design. The spectrograph is installed on rails which are mounted on a thermally isolated floor which allows the instrument to be opened. The vacuum vessel has a cut-away and reveals the internal optics (yellow) and detector systems at the rear. The total length of this system is ~6 m (Jurgenson et al. 2016)

.

be confident that we understand our detector at the sub milli-pixel level. Hence, the HARPS3 detector will be subject to a dedicated pixel mapping experiment that will measure the relative positions of every single pixel to less than 1 milli-pixel accuracy. This will result in a much more precise computation of the wavelength solution for each frame of spectral data and will contribute to minimising the uncertainty of each RV measurement.

HARPS3 is estimated to have a photon-limited RV error contribution of <1 m/s for all targets brighter than $V_{mag} = 9$ for a 20-minute exposure, see Figure 2.2.

2.1.2 The Isaac Newton Telescope

HARPS3 will be installed on the INT which will undergo a major refurbishment as part of the instrument installation. The motivation to automate observations comes from the need to increase efficiency over a 10 year period for the Terra Hunting Experiment. Instead of having a human-driven operational procedure, an automated scheduler will decide what observation is next in the list in real-time by factoring in target priority and local conditions, and will carry out the observations accordingly.

The INT has a primary mirror diameter of 2.54 m with an obscuration of 0.914 m in diameter resulting in a collecting area of 4.41 m^2. The INT has a median seeing of ~1.3″ which is anticipated to be reduced further with improved target tracking as part of the telescope refurbishments.

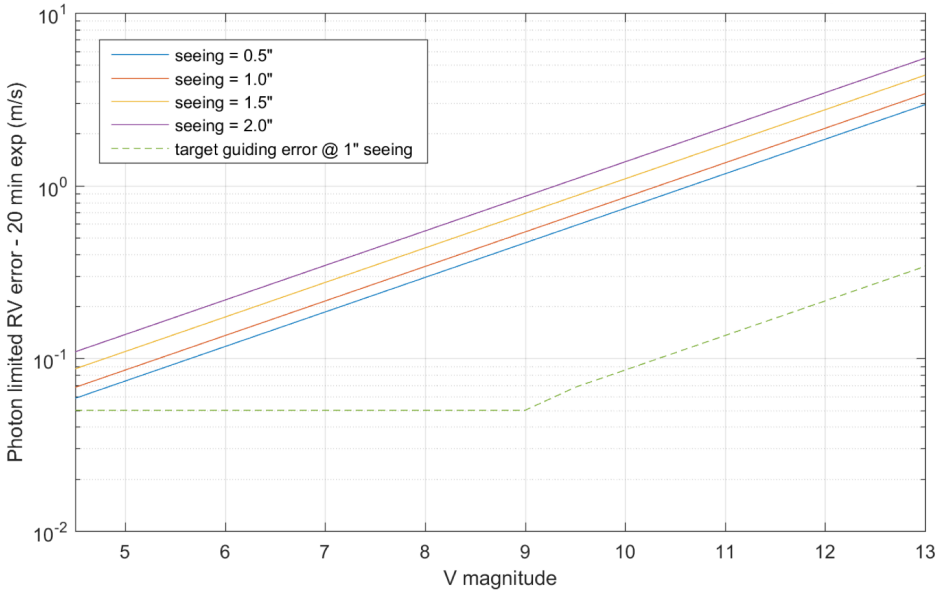


Figure 2.2: The predicted photon limited RV error for HARPS3 for a variety of seeing conditions and target magnitudes from Thompson et al. (2016). A typical $V_{mag} = 7$ target will exhibit 30 cm s^{-1} photon limited precision per measurement.

2.2 The Terra Hunting Experiment

The Terra Hunting Experiment has a single goal: to detect the first true Earth-twin candidate. This means a $1M_{\oplus}$ planet orbiting $1M_{\odot}$ star at a period of around 300 days. In this context, 'Earth-twin' contains no guarantee of habitability, follow-up observations would need to confirm the presence of an atmosphere. To accomplish this goal, it will observe a carefully selected set of ~50 targets every single night where possible for a total duration of 10 years.

To demonstrate the range of potential targets available to HARPS3 and the Terra Hunting Experiment, I have extracted a set of targets within 150pc from the GAIA database and filtered for stars with an effective temperature between 5000 K to 6000 K, Gaia Collaboration et al. (2016, 2018). This range limits the selection to G and K-type stars which could be similar to the Sun. Figure 2.3 shows this distribution plotted radially with the Sun at the centre.

The targets will be selected based on their magnitude, coordinates, and activity (more on this in Chapter 3), and a set of ~25 of them will be observed every night with a rolling rota throughout the calendar year. We anticipate that each target will be visible for 6 to 9 months before they are too close to the horizon and a different target becomes preferred for that observation date.

With such a schedule in place, and accounting for an estimate of cancelled observations due to local weather and periods of maintenance, the Terra Hunting Experiment will obtain over 1000 measurements for each target, and could be as high as 2000 for favourable targets.

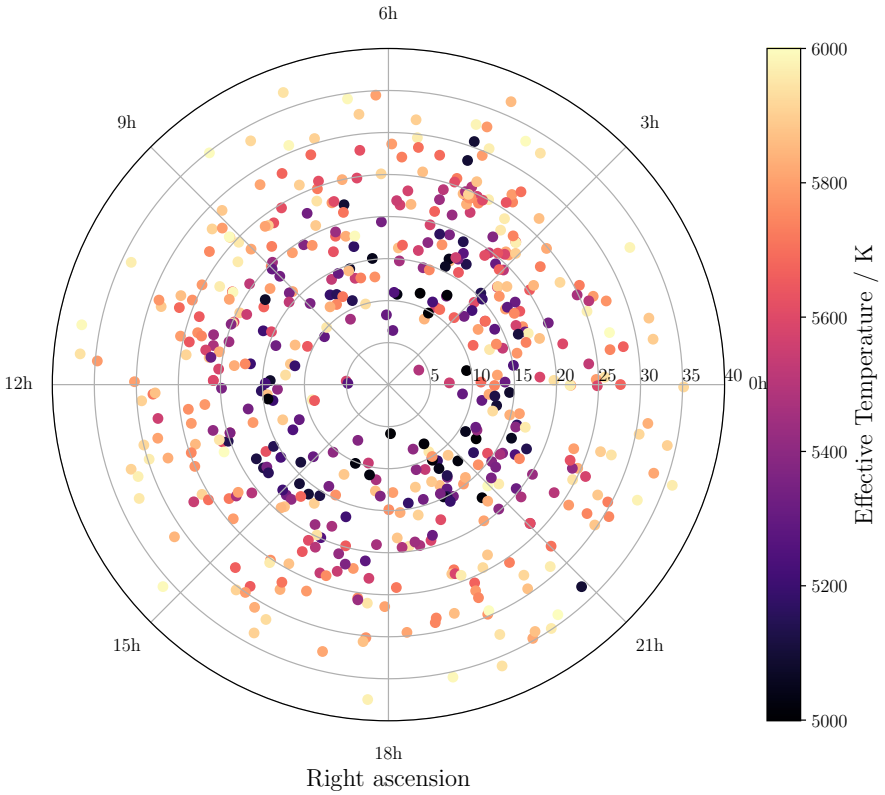


Figure 2.3: Potential nearby targets for the Terra Hunting Experiment measured by the GAIA instrument. These are targets filtered for their temperature which places them in the spectral classes of G and early-K stars. The apparent temperature dependence is a selection effect: it is easier to see hotter targets at a distance.

With this sheer volume of high-precision RV measurements per target, and a correlated set of polarimetry measurements, it should be possible to extract a 0.1 m/s signal should it exist as the complimentary data sets allow for precise tracking and mitigation of the stellar activity.

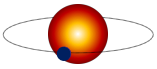
Chapter 3

Investigating the Feasibility of the Terra Hunting Experiment

'If you can dream it, you can do it.'

Enzo Ferrari

In this chapter I present a feasibility study of the Terra Hunting Experiment to assess the baseline capability of intense RV surveys for Earth-twin discoveries. This simulation is based on generating a set of RV time-series which contain a variety of: planetary systems, stellar signals, and photon noise, each time-series are sampled as per a typical schedule of the Terra Hunting Experiment along with some others. I used a multi-nested Bayesian sampling algorithm to assess the statistical likelihood of a Keplerian model of N planets fitted to the data. In this Bayesian framework, we are given full estimates of all parameter posteriors, and the Bayesian evidence of each model. Hence we can statistically assess the performance of each schedule in its ability to recover the underlying signal of the planets that are truly there. The simulation predicts that the Terra Hunting Experiment performs comparably to the continuous schedule despite only have $\sim 1/2$ of the number of data points. As this simulation effectively tries to ignore the stellar signals by brute-forcing a Keplerian fit, I conclude that there is a threshold where having better quality or more quantity of data does not yield better results without adequate stellar modelling as you are simply sampling the stellar oscillations with higher fidelity, hence we see only marginal improvement for the continuous schedule. This work was published as *On the Feasibility of Intense Radial Velocity Surveys for Earth-twin Discovery* (Hall et al. 2018).



3.1 Description of the problem

The aim of this investigation is to generate a set of RV time-series that represent a typical data-set from the Terra Hunting Experiment. As mentioned in Section 2.2, we anticipate that there could be up to 2000 measurements with 0.3 m s^{-1} photon-noise on a high-priority target, spanning a total of 10 years. This unprecedented level of high precision RV data has never before been captured on a single target, let alone a whole survey of nearby stars.

However, it is not the first time that long-term measurements of nearby quiet stars have been made with lower precision. Tau Ceti (HD 10700) has been a target of interest for Exoplanet studies as it is very similar to our Sun, exhibits low levels of stellar variability, and is only 3.7 parsecs away. Being an obvious candidate star for habitable planets, this star has been included in RV surveys since 1988 and it had been ruled out as hosting even a Jupiter-mass planet as recently as 2006 (Campbell et al. 1988; Paulson & Yelda 2006). However with new data this star was reported to host 5 low-mass planetary companions with the outer two straddling the habitable zone at 168 and 642 day orbits (Tuomi et al. 2013). Even though some of these planets are now thought to be aliases of the stellar oscillations (Feng et al. 2017), Tau Ceti proves to be a useful example of the need for better RV measurements over long baselines. I have plotted the data obtained from Tuomi et al. (2013)[a] in Figure 3.1 to show the vast RV spread even a quiet star can exhibit. Here, the average RV error of the data sets is 1.36 m s^{-1}, 0.54 m s^{-1}, 1.87 m s^{-1} for HIRES (High Resolution Echelle Spectrometer), HARPS and UCLES (University College London Echelle Spectrograph) respectively, ~2 to 5 times greater than the anticipated performance for HARPS3. In this plot I have also drawn a hypothetical super-earth in the habitable zone at 0.8AU simply for demonstration to compare the relative amplitudes of such a planet and the distribution of RV data from stellar activity and white noise.

This chapter is organised in 4 major parts. Part one, Section 3.2, discusses the shape and structure of the data generated by the simulation. Part two, Section 3.3, is an overview of stellar activity as the major source of false positives in RV surveys. Part three, Section 3.6, discusses the nested sampling algorithm used by this simulation to fit models to the data, and briefly introduces Bayes Theorem. Part four, Section 3.7, compares the results and discusses the performances of the different observation schedules used. At the end of the chapter I discuss potential avenues that this simulation could go should it be extended and what my immediate next steps would be for follow-up work.

[a] For this plot the HARPS data was converted from km s^{-1} into m s^{-1} and given a mean offset of 0 m s^{-1}

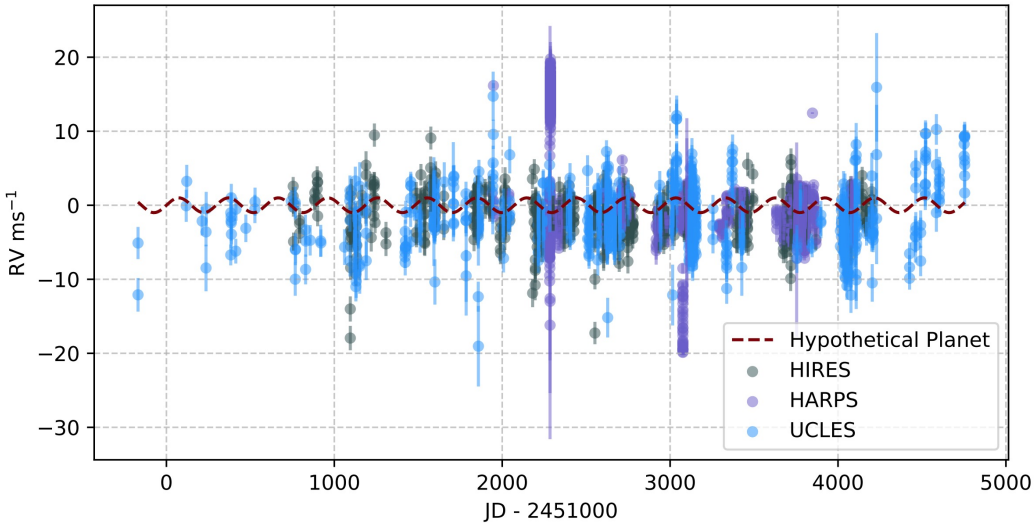


Figure 3.1: Almost 14 years and over 5000 measurements of RV data has been captured on star Tau Ceti with the HIRES (grey), HARPS (cyan), and UCLES (blue) instruments. I have superimposed the RV from a *hypothetical* super-earth orbiting at 0.8AU simply for demonstration, and filtered the data-set to omit any $|RV| > 20\,\mathrm{m\,s^{-1}}$ outliers for clarity.

3.2 Simulating RVs from a Terra Hunting Experiment Perspective

To generate a typical radial velocity time series that we expect from the Terra Hunting Experiment, I used the following procedure to create a database of RVs which would eventually be fed into the nested sampling algorithm for model comparison.

1) Define a set of solar systems (Table 3.1)
2) Generate planets each with Keplerian parameters per system
3) Define a set of time-stamps from various schedules
4) Factor in predicted seasonal weather effects and generate the Keplerian RVs
5) Add white noise drawn from a $1\sigma = 30\,\mathrm{cm\,s^{-1}}$ normal distribution
6) Add RVs from a stellar model (SOAP2.0)
7) Combine for a set of RVs for each solar system and observation schedule

3.2.1 Defining Solar Systems for the study

For items 1 and 2 in List 3.2 I used the formulation discussed in Section 1.2, specifically Equations 1.8 and 1.10. For practicality, Equation 1.8 was coded as

$$v(t_n, j) = V_j - \sum_{p=1}^{N_p} K_p[\cos(\frac{2\pi}{P_p}t_n + \theta) + e_p \cos\omega], \tag{3.1}$$

where V_j is the systemic velocity relative to the observer, $v(t_n, j)$ is the total system velocity including the contribution from N_p planets, and P_p is the period of the p^{th} planet, and t_p are the time-stamps at which the RV of the system was measured.

I knew, a priori, that I was going to fit models to these datasets with a nested sampling algorithm whose analysis time scaled with the number of parameters considered so at an early stage in the investigation I decided to simplify the study. I forced all of my planets onto circular orbits, i.e. eccentricity = 0, I set the system inclination to zero in all cases such that $\sin i = 1$, and I eliminated the systemic velocity relative to the observer by setting $V_j = 0$. This adjustment simplified Equations 1.10 and 3.1 to

$$K = \left(\frac{2\pi G}{P}\right)^{1/3} \frac{M_p}{(M_\star + M_p)^{2/3}}., \tag{3.2}$$

and

$$v(t_n, j) = \sum_{p=1}^{N_p} K_p[\cos(\frac{2\pi}{P_p} t_n + \theta_p)]. \tag{3.3}$$

In this simplified regime, we have just three parameters per planet to fit (period, mass, orbital phase) as opposed to 5 parameters per planet plus the systemic velocity of the system. As our system velocity is simply a co-added series of independent planetary RVs, I have assumed that there are no significant planetary interactions. I justify that this is the case by considering the strongest planetary interaction of all my solar systems - a long-period gas giant like Jupiter in the same system as an Earth-twin. I calculated that when these two planets are closest together, the gravitational force of the gas giant imparts a maximum radial velocity on the Earth-twin of just $0.000\,04\,\mathrm{m\,s^{-1}}$ due to it being almost 4AU away from our Earth-twin at closest approach, an amount negligible for this study. It is also generally thought that low eccentricity is desirable for habitable planets to maintain gentle seasonal variations (Bolmont et al. 2016; Wang et al. 2017), something which has greatly benefited life on Earth which has an eccentricity of just 0.0167.

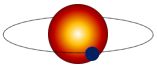
Also, as the Terra Hunting Experiment is a precursor for an Earth-twin habitability study, I have assumed that my solar systems are stable over long timescales (many hundreds of millions of years), and that the planets orbits are effectively fixed. There are many studies on planetary dynamics, early system evolution, and the formation of stable solar systems, and various published code-bases where a user can test the stability and look at the long term evolution of a set of planets. I elected to not check the stability of my systems as for this feasibility study, I simply wanted to generate signals that were *representative* of the types that the Terra Hunting Experiment was trying to find. I.e. can I create a series of radial velocities that are sampled appropriately, contain stellar oscillations, sinusoidal planetary signals, and measurement white noise? If so then I do not need to worry about whether these hypothetical

Table 3.1: A list of planetary systems modelled in the RV simulation. System 4 is the null–case and contains no planets. For each case, the star has a mass of $0.8 M_{\odot}$.

System	Planet Mass / $M_{\oplus}$	Period / days	RV / m s^{-1}
1	1.00	293	0.11
2	0.82	197	0.11
	1.00	293	0.11
	200.00	2953	10.34
3	1.00	101	0.16
	1.00	197	0.13
	1.00	293	0.11
4	N/A	N/A	N/A

systems are dynamically stable or not over millions of years. However, as I will discuss in more detail in Section 3.11, a full yield estimator of the Terra Hunting Experiment *would* need to consider dynamically stable systems as such a simulation would likely generate many thousands of systems for some statistically rigorous results. For such a study, any large-scale assumptions about the solar systems can greatly skew the results.

With the formulation of the RV semi amplitude in place, I created a set of 4 solar systems to act as test cases for the study, see Table 3.1. Each system was chosen to test the analysis in a different way and whilst they contain different planets, the central star is the same in each case. The first system is a simple lone Earth-twin orbiting at a period of 293 days. This is a simple test case of whether or not the analysis pipeline can find a 0.11 m s^{-1} signal. The second system is an analogue of our own Solar System but only including three contributions to the total RV of the star. In this system there is the same Earth-twin, but it now accompanied by a Jupiter-analogue and a Venus-analogue. This system acts as a test of whether the analysis can find the Earth-twin when there is a large and longer period RV component from the gas giant, and when there is a similar signal but on a shorter period. The third system is three Earth-twins orbiting at different periods. This is a test of our analysis to differentiate similar amplitude signals on different periods - a test of distinguishing more complex signals. Lastly, the fourth system contains no planets and is our null hypothesis case. Here we want to examine what, if any, false positives our analysis finds in this data when there is only white noise and stellar oscillations. The four systems are summarised in Table 3.1. The reader may have noticed the obscure periods at which each planet is placed. These numbers arise from scaling down the Earth, Venus, and Jupiter orbits for a $0.8 M_{\odot}$ star, and then rounding to the nearest prime number. Having prime numbers for our analysis gives us confidence in a detection when we anticipate many aliases of a ground-based schedule that has 10, 30, 180, and 365 day repeating observation windows.



3.2.2 Defining Observation Schedules and Dealing with Weather

For point 3 in List 3.2, I created two other schedule categories to act as a benchmark for comparison to the Terra Hunting schedule. One is a schedule used in current ground-based RV surveys on HARPSN (an ideal comparison but on an oversubscribed instrument), and a continuous schedule with an observation every day for the entire 10 year survey.

As the Terra Hunting Experiment target list is currently not finalised, it is difficult to say exactly how the 10 years of observations will be distributed on a night-to-night basis. At the time of conducting the study, the target list had yet to be narrowed down so I created a as-generic-as-possible schedule to stand in its place. For this survey we know that we want to observe every star every night where possible, and that the target list will hopefully have an even distribution of targets throughout the calendar year. Thus for this schedule I generated a list of observations that span for 6 months, and then have a 6 month gap, and repeats for 10 years. Here, the gap represents another target taking priority in the other six months of the calendar year. Whilst unrealistic, each observation is conducted at exactly midnight. This is due to a lack of a finalised target list so it was impossible to distribute observations throughout a given night with only 1 imaginary target, and I would be injecting a bias/systematic if I *decided* where I thought best to place the observation. The only other sensible choice would be to pick a slot in the night drawn from a random distribution. But this is not how an actual survey would decide the observations, at least not one trying to observe the sky efficiently, and I can handle there being a 24 hour alias in the data as I am only interested in our ability to detect signals at many hundreds of days.

For the reference schedule, I emulated a schedule that may be obtainable on an instrument as similar as possible to HARPS3. The purpose of this is to see if it is feasible to conduct the Terra Hunting Experiment on an existing telescope without the need for building HARPS3. Here I refer to Dumusque et al. (2011a) where the authors discuss the relative strengths and weaknesses of distributing available observations throughout a season and assess the results. In their comparison they noted that the HARPS-GTO program could obtain one measurement per night for 10 consecutive nights per month for a duration of four years (Lovis et al. 2006a). I used this paradigm and created a schedule called the ‘reference’ schedule as it is the schedule to beat for the Terra Hunting Experiment. As with the previous schedule, I place these observations at exactly midnight of each night and only include the first six months of the year.

The last schedule category is an idealised hypothetical schedule I refer to as the ‘baseline’. It represents either HARPS3 in space, or a network of identical instruments located around the world that are all part of the survey, or a suitable target that may have this coverage. For this schedule I simply allocate an observation at midnight, every single night for the entire 10 year

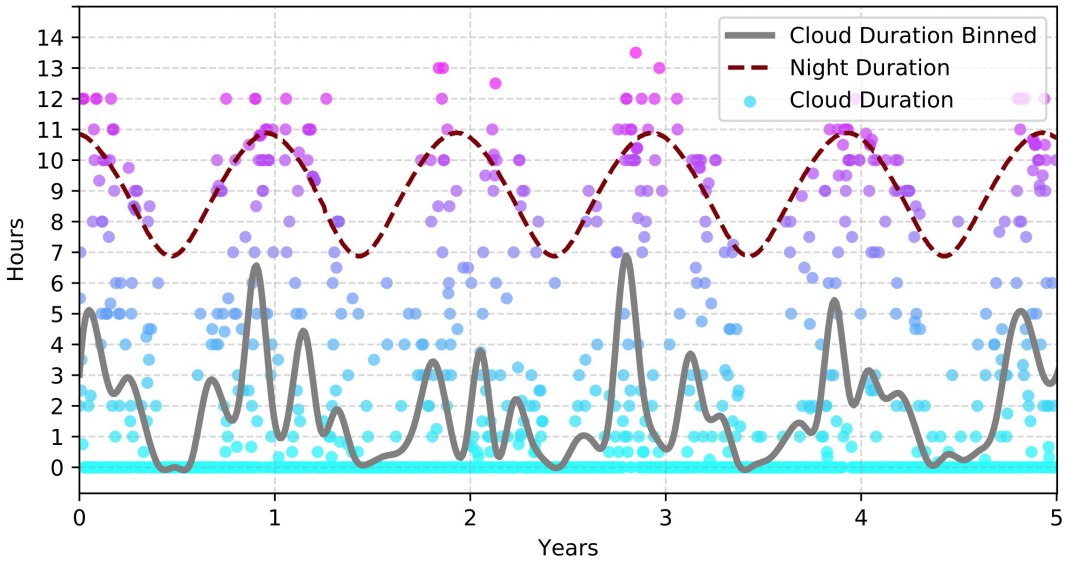


Figure 3.2: The measured duration of cloud cover per night from La Palma (scatter). A rolling mean of 30 day bin is shown (solid line) along with the hours of darkness excluding twilight (dashed line). 'Year 0' is 2010 and each of the year tick-marks are placed on January 1st.

survey.

For the ground-based surveys the seasonal weather must also be considered, point 4 in List 3.2. As HARPS3 will be situated in the northern hemisphere, we anticipate that the summer months will have less inclement weather than the winter and so our hypothetical schedules would yield different results for two stars that are observed in the summer versus the winter. The Terra Hunting Experiment consortium has access to historical weather logs from La Palma which I can feed into my simulation in order to decide whether or not an observation is taken.

A snapshot of the weather data is displayed in Figure 3.2. The yearly variations are apparent in the correlation between the length of night and the amount of cloud cover, and the 30-day binned data shows spikes in the winter months around the year tick-marks (centred on 1st January). For the each observation in the simulation, I chose to cancel it if the duration of cloud is greater than half the duration of the night. I.e, in a normally distributed set of observations throughout a long baseline (many years), I will cancel the observation if there is a $> 50\%$ chance it falls on a cloudy section of night. I have to take this statistical approach because the original database of weather information the data is 'minutes of cloud per night', not the actual times of the cloud coverage itself. However, averaged over a sufficiently long timescale and with enough data-points, this is probably a good representation of how weather seasonally affects an observation schedule. To see the summary of how this affects the various schedules, see Table 3.2. The 'Season' column refers to when I have centred the 6-month window for

Table 3.2: A list of observation schedules used in the simulation with the total number of observations possible for each schedule per season after considerations for the weather have been taken.

Survey	Season	Duration / years	Total Observations
THE	Summer	10.0	1695
	Summer	5.0	847
Reference	Summer	10.0	507
	Summer	5.0	259
Space	All-year	10.0	3650
	All-year	5.0	1825

the ground based schedules. To maximise the number of observations, the season runs from March-October of each calendar year.

For point 5 in List 3.2, I added photon noise to each measurement drawn from a normal distribution of $1\sigma = 30\,\mathrm{cm\,s^{-1}}$. This value represents the photon limited shot noise on each measurement in optimistic seeing conditions, shown in Figure 2.2, and was calculated with a 20 minute exposure of a $V_{mag} = 7.5$ target.

Lastly, I include the RV variations from the physical processes of the star itself. It is well known that stellar oscillations are a limiting factor in low-mass planet detection, and that modelling and mitigating these effects is key for Earth-twin discoveries. Some recent articles on stellar activity, modelling, summaries, and how we can use the Sun as a distant star to aid our understanding include Dumusque et al. (2011a,b, 2015c); Herrero et al. (2016); Fischer et al. (2016); Haywood et al. (2016). Before I discuss how I integrate stellar oscillations into the simulation, I first must discuss in brief the various stellar physical processes that give rise to RV variations.

3.3 Stellar activity as a source of false positives

In this section I outline the key physical processes that give rise to RV variations from stellar activity. An excellent recent summary of these, complete with many additional details and references, can be found in Chapters 1 and 4 of the thesis of Haywood (2015).

Stellar signals manifest in the data as a source of quasi-periodic noise on varying temporal baselines that can be falsely interpreted as planetary signals (false positives) (see Queloz et al. 2001; Huélamo et al. 2008; Haywood et al. 2014; Robertson et al. 2014; Rajpaul et al. 2016 for examples of previously claimed planets being debunked as stellar activity). Stellar activity can be classified by four physical phenomenon, each with a characteristic RV amplitude and period, I list them below and then give a brief overview of each.

1) Oscillations from pressure waves propagating in the convective zone of the star; RVs of $0.1\,\mathrm{m\,s^{-1}}$ to $4\,\mathrm{m\,s^{-1}}$ over a few minutes (Dumusque et al. 2011a).
2) Granulation from convective cells at the stellar surface where cool material falls into the star whilst hot material rises from beneath in small cells, RVs of $0.1\,\mathrm{m\,s^{-1}}$ to $4.0\,\mathrm{m\,s^{-1}}$ over minutes to days (Dumusque et al. 2011a).
3) Short-term stellar activity from rotating surface structures as the star rotates, e.g. sunspots creating an RV asymmetry as they move from the red to the blue-shifted half of the star; RVs of $0.5\,\mathrm{m\,s^{-1}}$ over 20 days (Lagrange et al. 2011).
4) Long-term magnetic cycles that drive spot formation rates, rising granules of plasma, and overall brightness; RVs of a few $\mathrm{m\,s^{-1}}$ over many years (Meunier & Lagrange 2013).

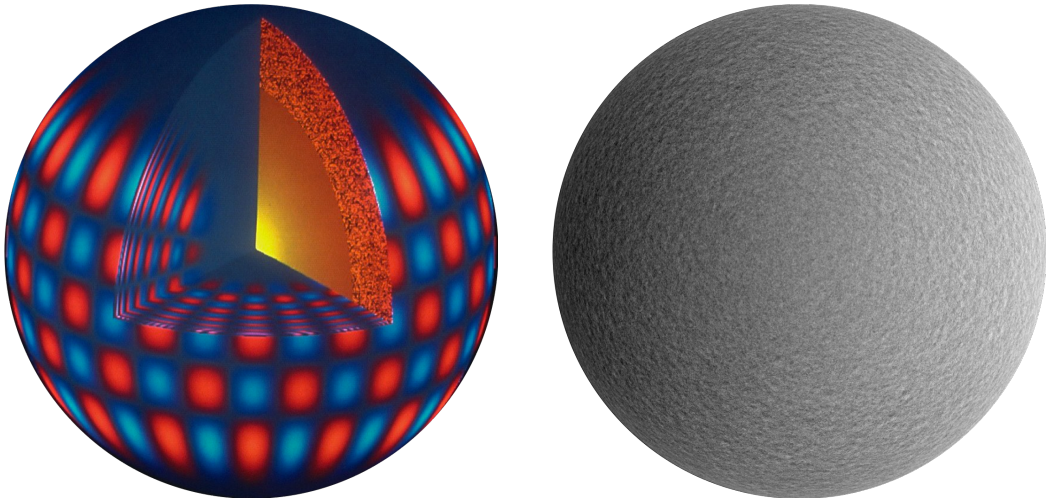
3.3.1 Oscillations

The internal structure of stars is constantly oscillating. Small pressure variations move through the material and result in physical deviations on the surface where they can be modelled as the oscillations of a sphere of liquid in hydrostatic equilibrium. Depending on the mode of oscillation, different cells on the surface will have a velocity perpendicular to the static surface of a sphere, see Figure 3.3a. Many different modes can co-exist simultaneously and the observer will see the superposition of all of them. Yet, coherent oscillations are observed on the sun, and other stars, to have periods of a few minutes (~mHz in frequency) and amplitudes of around $1\,\mathrm{m\,s^{-1}}$ to $5\,\mathrm{m\,s^{-1}}$. To help mitigate these effects, observers typically take long exposures of 15 to 20 minutes as this averages out most p-mode oscillations, (Chaplin et al. 2019). In Dumusque et al. (2011a) the authors used synthetic data to determine an ideal strategy by comparing different sets of exposures in a given night (1×30 min or 3×10 min for example). They showed that multiple exposures per night, repeated daily for binning over $5-10$ days where possible, reduced the RMS of their RV measurements by up to a factor of 4 (from $1.6\,\mathrm{m\,s^{-1}}$ to $0.4\,\mathrm{m\,s^{-1}}$ for the case of μ Ara).

To better understand these effects, we are fortunate that we have a star nearby which we can fully resolve - the Sun. The Sun is unique in this case as it is the *only* star whose surface can be sufficiently resolved to reveal detailed structure. By combining sun-as-a-star measurements (Dumusque et al. 2015c) with fully resolved images of the solar disk courtesy of the Solar Dynamic Observatory (Pesnell et al. 2012; Haywood et al. 2016), we can correlate radial velocity measurements with physical features, structures, and active regions.

3.3.2 Granulation

Granulation is the name given to the phenomenon of hot material rising in the convective zone and subsequently falling as it cools. To an observer, the granules appear as bright cells (hotter



(a) A single helioseismology p-mode model (b) The line-of-sight velocity of the sun

Figure 3.3: a) An example of a single p-mode of pressure waves propagating through the Sun. The red cells are redshifted and moving away from the observer, whilst the blue cells are blueshifted as they are moving towards the observer. In the radial direction towards the core, the wavelength increases due to the increasing speed of sound of the higher density material. And b) the observations of the line-of-sight velocity of the Sun as observed by the SDO. The more chaotic pattern is due to the superposition of many modes simultaneously. The darker right half and brighter left half of the disc is due to the solar rotation.

rising regions) with dark boundaries (cooler falling regions). They are typically 1000 km in size and last around 10 min. A single granule will have an RV of $2\,\mathrm{km\,s^{-1}}$, but there can be upwards of one million present on the stellar disk at a time. Hence, the total effect can be effectively binned down to $\sqrt{N}$ and so contribute $\sim 2\,\mathrm{m\,s^{-1}}$ to the RV of the Sun. As with p-mode oscillations, exposures of 20 min help mitigate this effect further. In Figure 3.4, granules and a sunspot are shown with the Earth to scale, the image is from the Swedish Solar Telescope[b].

3.3.3 Rotating features: Spots, Faculae, and Plage

The most problematic stellar activity for planet hunting is in the effects of rotating surface features. The solar rotation rate ranges from 24 to 28 days from the equator to the poles respectively, so any surface feature on the sun will induce an RV variation on that timescale and can easily be misconstrued as a planetary RV as mentioned in Section 3.3.

Sunspots are caused by local intense magnetic flux loops that emerge from the solar surface. Here, the strong magnetic fields inhibit the convective cycles that give rise to granulation and cause a visibly darker and also cooler (600 K to 1800 K less than the photosphere, Solanki 2003) region. As they inhibit the rapidly rising and blueshifted convective cells, a single sunspot will

[b]Image data obtained from https://www.isf.astro.su.se/

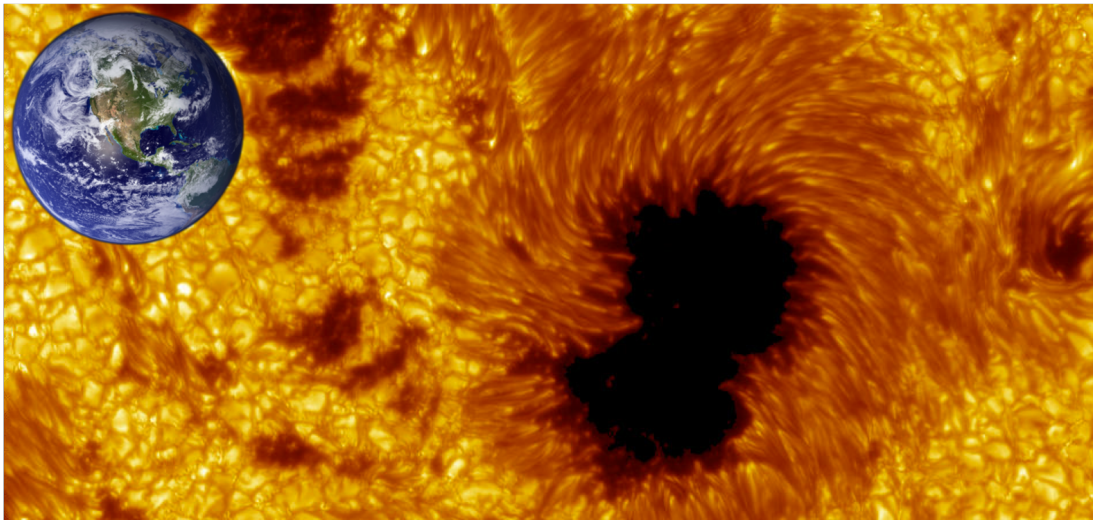


Figure 3.4: An example of the cellular structure of granulation accompanied by a large sunspot. The Earth (~13 000 km diameter) is drawn to scale.

cause an RV variation as it moves across an otherwise featureless disc where the equally red and blue-shifted sides usually average out.

The lifetime of a sunspot is proportional to its size, smaller spots can live for just days whilst the largest live for many hundreds of days. They decay when the magnetic flux loops decrease in density, and they slowly diffuse into the photosphere. Hence a larger spot with a higher area-to-perimeter ratio will take longer to decay (Berdyugina 2005).

On the sun, sunspots form in restricted regions of latitudes (±35° for the sun, can be higher for others), and also tend to reoccur in specific longitudes known as active longitudes in localised areas of increased magnetic activity.

Aside from spots, there is another other surface feature that greatly affect the RV of a star: the faculae. Faculae are bright, thin tubes that fill the intergranular lanes between granules surrounding spots, but they can also exist in isolation. They are only 100 K hotter than the photosphere, so head on they have relatively low contrast. Towards the limb however they are more visible, see Figure 3.5[c]. Despite not having a large contribution to the photometric variability of stars, they do contribute strongly to spectroscopic variations as they are often associated with the suppression of convective blueshifted regions around sunspots. Faculae have short lifetimes of under an hour. Plage are photometrically bright regions constructed from small bright points known as flocculli, plage are the chromospheric counterpart of the faculae. Like faculae, they are most visible nearer the limb and have very short lifetimes.

There is no simple method to mitigate stellar rotation effects as there is with granulation or

[c]Frames captured from a video in the SDO gallary at https://sdo.gsfc.nasa.gov/gallery/main/item/569.

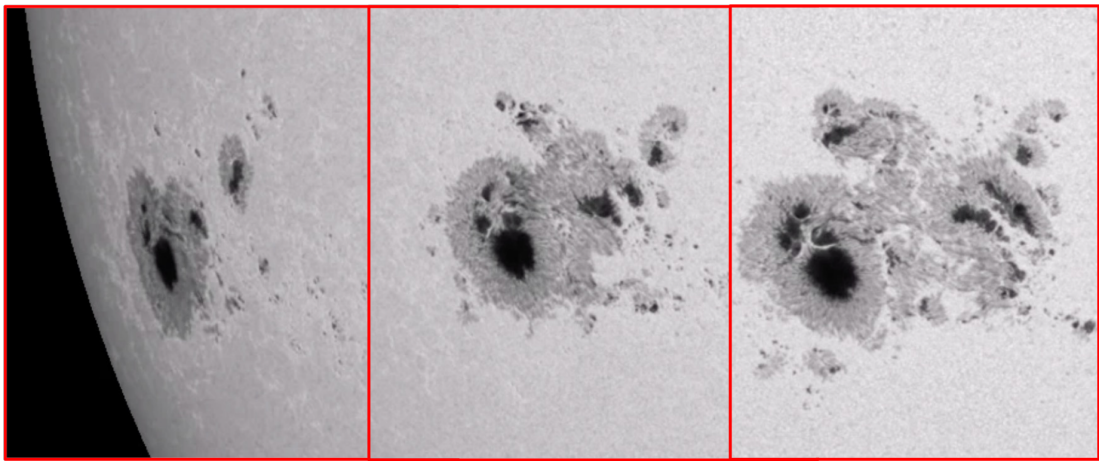


Figure 3.5: A spot group evolving as it cross from the limb to (left frame) to the centre of the disc (right frame). The visibility of the faculae (bright filaments) decreases as the group moves away from the limb as we no longer look at the sidewalls of the structure.

pressure waves such as a long exposure. Their effects on the RV curve last for many days or even weeks, are quasi-periodic, and are hard to predict. There are a number of *indicators* we can use to assess the influence of rotating features, and to distinguish them from the Keplerian signals of planets. I discuss a few of these indicators in more detail in Section 3.5.

3.3.4 Magnetic Cycles

Over longer timescales, the overall activity of a star can fluctuate. The sun has an 11-year activity cycle, as shown in Figure 3.6, which dominates the number and position of spots, and the occurrence rates of coronal mass ejections. At the peak of a solar cycle, associated with high magnetic field strengths, the sun may have up to 300 spots visible whilst at the minimum of the cycle the number can be zero. In the context of planet hunting, more spots and activity generally means larger RV variations and a higher RV RMS on each measurement which greatly inhibits small amplitude signal detection. Figure 3.6 also demonstrates how the positions of the active regions migrate towards the stellar equator. Not all solar cycles are identical, there are hypothesised cycles of many hundreds to thousands of years which could account for the differences in spot numbers between each cycle (Damon & Jirikowic 1992).

3.4 Including Stellar Activity in the Simulation

With a description of the various processes that lead to RV variations from an active star, I can now describe how I incorporated this variability into my datasets.

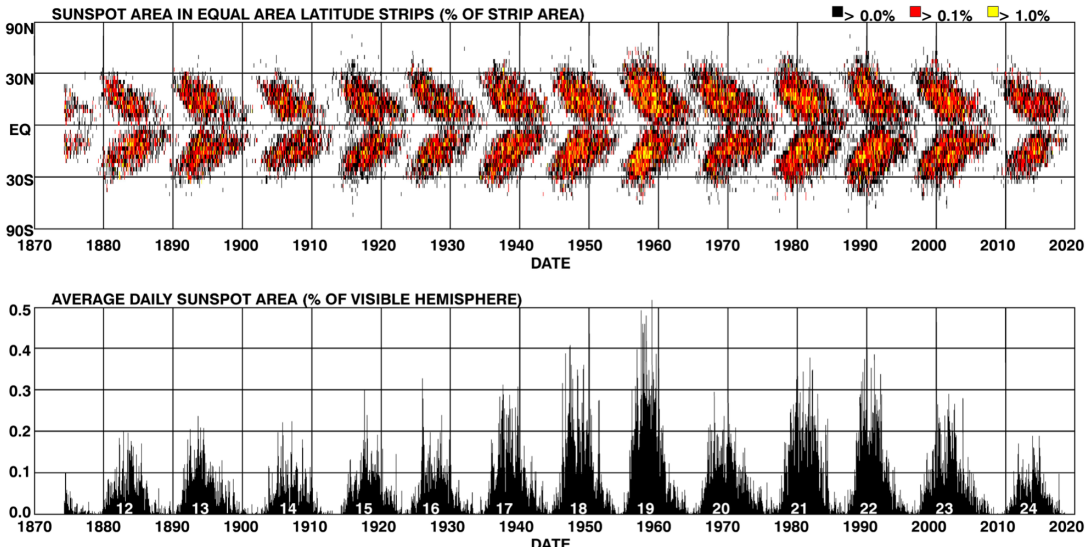


Figure 3.6: The butterfly diagram of spot location and area coverage (top) and the average daily sunspot area (bottom) from 1875 to 2019, Hathaway (2019). Each 11-year cycle has a characteristic butterfly-wing shape as the spots form closer to the equator during the cycle, hence the name.

3.4.1 Using SOAP2.0 to Compute Active Region RVs

For the effect of rotating surface features, I use the implementation provided by SOAP2.0 (Stellar Oscillation and Planet, 'SOAP' from here on) (Dumusque et al. 2014, 2015a). The SOAP code considers the visible half of star as comprising of a large number of equal area cells[d]. In each cell, the velocity is derived from the stellar radius and rotation speed at that cell. SOAP calculates the RV of each cell by considering the CCF (see Section 1.2.1) at the cells coordinates that is modulated by a quadratic limb-darkening law. The RV of all the cells is then integrated to a single CCF and a Gaussian is fitted to obtain an RV value,

$$CCF_{quiet} = \sum_{x,y}^{N} I_{LD}(x,y) CCF_{x,y}. \tag{3.4}$$

Where x, y are the coordinates of a particular cell, N is the total number of cells, $I_{LD}(x,y)$ is the limb-darkening term for that cell, and $CCF_{x,y}$ is the CCF of that cell. For a completely featureless star, the RV averages to zero.

Should an active region, such as a spot or faculae, exist on the surface it will change the weight CCF of the cells affected by the same quadratic limb-darkening law, but its Gaussian

[d]Equal area in projection as seen by the observer, not equal area on the surface of the star.

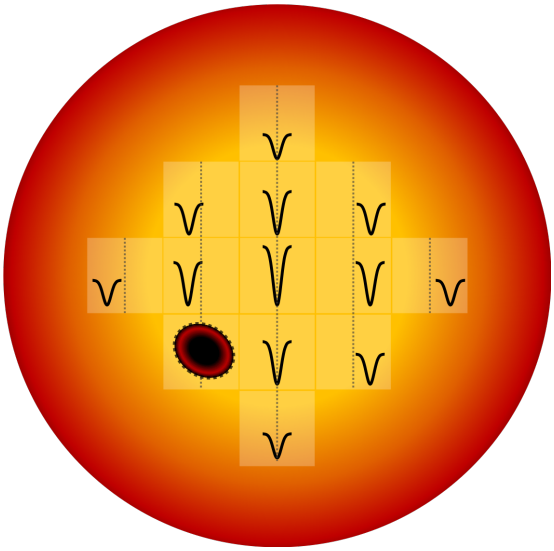


Figure 3.7: An example of how SOAP breaks up the stellar disc into a discrete set of cells, and calculates the RV of the individual cell, $0\,\mathrm{m\,s^{-1}}$ is represented by the dotted line down the middle of each cell. Towards the edge the CCF is weighted due to the limb-darkening effect. The star is rotating such that the left half is blueshifted towards the observer and the right half redshifted away. A large spot is in the lower-left cell and completely occupies it so no flux is observed. In practicality, SOAP uses $\sim 10^5$ cells.

shape will remain. Here, a difference to the CCF from the affected cells is given by

$$\Delta CCF = CCF_{active} = \sum_{x,y} I_{LD}(x_a, y_a)\left[CCF_{x_a,y_a} - I_a(x_a, y_a)CCF_{x_a,y_a}\right], \tag{3.5}$$

where I_a is the ratio of the Plank function evaluated in the active region and the continuum of the star.

For the final CCF of the star, CCF_{quiet} - CCF_{active} is computed. Everything described so far returns a single RV measurement of a stationary active region at a single point in time. With SOAP the user can enter a list of time-stamps at which to calculate the total CCF. SOAP will rotate the star by the sufficient amount between time-steps and return the new RV measurement with the active region now on a different part of the stellar surface. Should the active region pass to the reverse side of the star, the CCF will have a value of $0\,\mathrm{m\,s^{-1}}$

As discussed in Section 3.3.3 and showed in Figure 3.5, spots, spot groups and faculae all evolve over time. The RV due to an unchanging rotating spot group would result in a pure sinusoidal-like signal. SOAP addresses this by allowing the spots and spot groups to dynamically evolve as they rotate around the surface.

The user can control a few parameters such as average spot lifetime, average size of spots, average number of spots per group, how many groups and where they are likely to form, and the effective stellar temperature to spot temperature ratio. Observations of the Sun from various

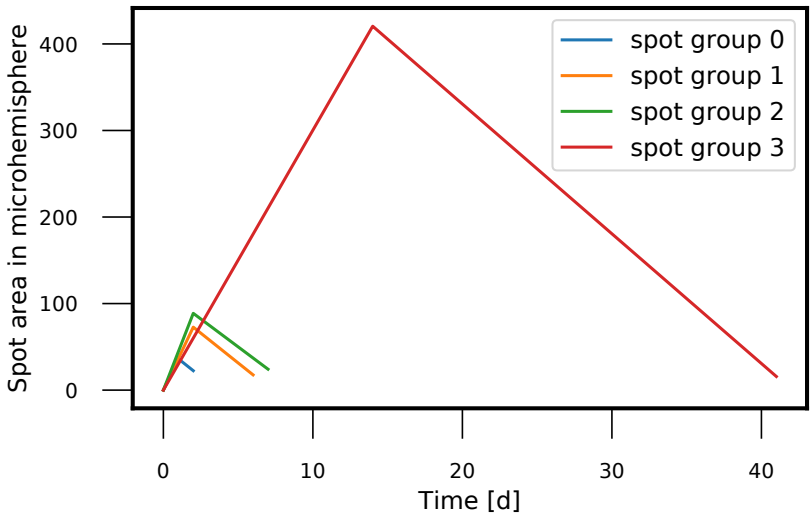
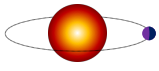


Figure 3.8: An example of the area of one set of spot groups as a function of their lifetime. Each group occupies a unique location on the surface of the star and is independent of the other groups.

instruments (SDO, SST, HARPSN Solar Telescope etc) tell us that spot groups on the Sun live from between a few hours to a couple of months. For our implementation of SOAP, I follow the default settings for a quiet sun, and generate our spot group lifetimes such that 50% of spot groups have lifetimes between 1h and 2 days, 40% of spot groups have lifetimes between 2 and 11 days and, 10% of spot groups have lifetimes between 11 and 60 days. I also follow the default settings and create 4 spot groups in total. Each groups' lifetime is drawn from a random distribution weighted to the previously mentioned percentage chances. In Figure 3.8, the surface area of four spot groups are shown as a function of time. Their size is determined by SOAP forcing them to grow for the first ~1/3 of their life, and then shrink to 0 for the remaining ~2/3 of their life. Once a spot group has disappeared, SOAP will regenerate a new group whose parameters (size, lifetime, number of spots) are drawn from normal distributions. For visualisation purposes only, SOAP generates a video of the spot groups rotating on the surface. In Figure 3.9, 5 frames have been captured to demonstrate the rotation and evolution of spot groups and placed side by side to show the slight differences at each time step. For this visualisation I generated an active Sun with around 100 spots at a given time and 4 clear spot groups placed 90° apart in alternating latitudes above and below the stellar equator. Each frame is ~60 h apart, and the total sequence covers almost one half of a stellar rotation.



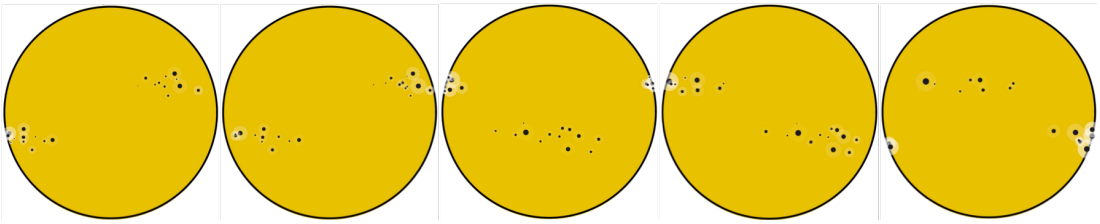


Figure 3.9: Five frames captured from the results of SOAP. Two spot groups can be seen crossing the stellar disk, and the spots evolve during the 12 day sequence. The brighter halos around each spot near the limb are the representation of the plage becoming more visible. The limb darkening of the disc is taken into consideration but is not shown in the representation. This sequence is comparable to that shown in Figure 3.5.

3.4.2 SOAP with the Terra Hunting Experiment

For the Terra Hunting Experiment I used SOAP to generate 10 years of RVs representative of a quiet, 0.8 $M_{\odot}$, P_{rot} = 25 d, Sun-like star. I set the average spot count to 20, randomly distributed through 4 spot groups. The number of spots was chosen by examining the activity of our Sun over the past 80 years and noticing that this value represents the activity of the Sun during its quiet periods, see Figure 3.10[e]. I then generate 10 years of RVs with a measurement spaced 24 h apart. Ideally, I would generate a near continuous RV series with which I can sample arbitrarily for my given schedule, but SOAP is CPU intensive due to the fact it has to integrate the entire CCF, fit a Gaussian, and extract the RV (along with other parameters such as the BIS and FHWM, defined in Section 3.5), evolve the spots, dynamically generate new spots, and then rotate the star at each time-step. On a single core at 2.5 GHz, it would take well over 100 hours to generate the full 10-year data set. This limitation was also one of the influences to force the observation schedules to a 24 h cadence. Should I want a schedule that has time-stamps on a 20 min cadence, I would need to generate 72 times the SOAP data, which would take upwards of 300 days. I discuss this and other limitations of SOAP in Section 3.4.4.

In Figure 3.11 I show a 400 day zoom from the full RV series of our quiet star. The RMS of the RV is ~4 m s^{-1}, 40 times greater than the signal of an Earth-twin. The chaotic and quasi-periodic pattern is due to a number of spots in different spot groups evolving as the star rotates. This was the desirable outcome for using SOAP to create a major source of false positives.

Returning back to list in section 3.2, we are now in a position to combine everything into RV series that span a 10 year baseline and are sampled as per the defined schedules. In Figure 3.12, the 10 RV series from the three schedules observing the Solar System analogue of 'System 2' (see Table 3.1) are shown. Each time-series has the Keplerian signals from the three planets,

[e]Data obtained from http://sidc.be/silso/home.

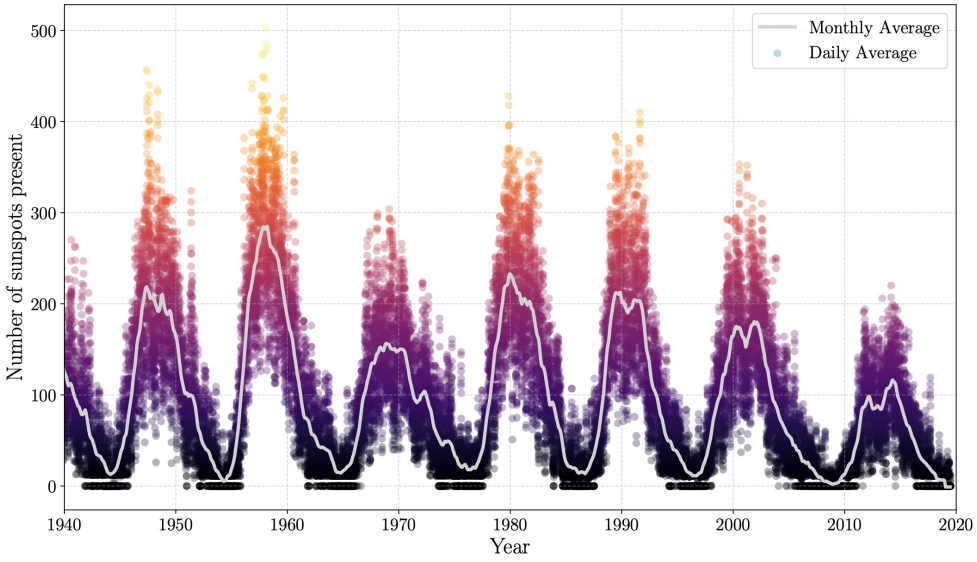


Figure 3.10: The daily and monthly-averaged numbers of spots present on the sun over the past 80 years. This data was used to justify the choice of 20 spots from SOAP to represent the activity levels of a sun-like star during a low-activity phase.

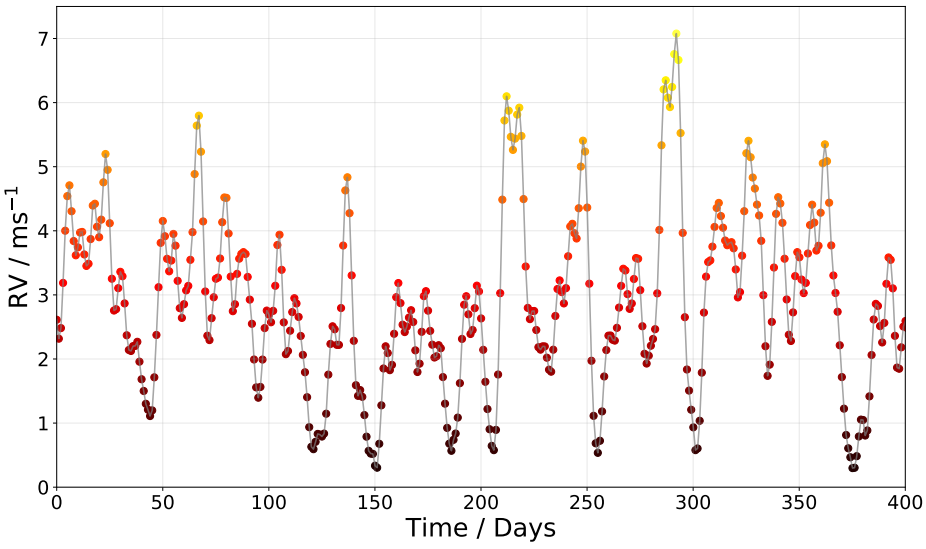


Figure 3.11: An example of the raw RV series from surface features on a quiet solar-type star generated from SOAP2.0. As SOAP returns discreet points every 24 hours we plot a smooth interpolation to better show the structure of the data.

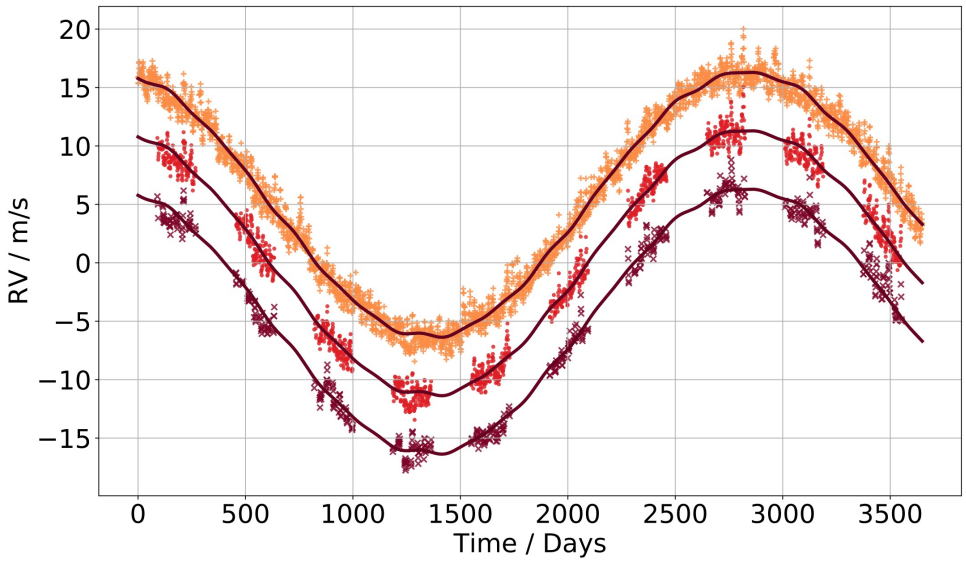


Figure 3.12: A 10–year simulation of the RV series of System 2 (see Table 3.1) including SOAP2.0 and Gaussian noise, observed with a typical Terra Hunting summer schedule (middle), Reference summer schedule (bottom), and the uninterrupted/space schedule (top). The data sets are staggered vertically for clarity, and the solid lines through each are the pure RV curves from the Keplerian planetary models.

photon shot noise drawn from a distribution of $\sigma = 30\,\mathrm{cm\,s^{-1}}$, and the 10 year SOAP data from a quiet star. The large sinusoidal signal is the $12\,\mathrm{m\,s^{-1}}$ Jupiter-like planet dominating the radial velocities.

Figure 3.13, a zoomed section of Figure 3.12, better shows the scale of the RVs from SOAP compared to the slight bump from one of the Earth-twins in this system. The space and Terra Hunting schedules have sufficient data density to show a large spot group crossing the stellar disc at around 1275 days, whilst the reference schedule only captures part of it. It is clear that the stellar signals need to be accounted for to high fidelity for all schedules in order to recover Earth, and the better sampled schedules allow the reconstruction of these different signals (activity + planet) with much less degeneracy whilst the lesser sampled schedules will likely struggle.

The final data-set comprises of 4 planetary systems, observed by 6 schedules, with the stellar noise either present or missing. This totals 48 RV time-series to analyse for the presence of planets. How well the true planets are recovered is expected to be a function of both the data density and the presence of the stellar signals.

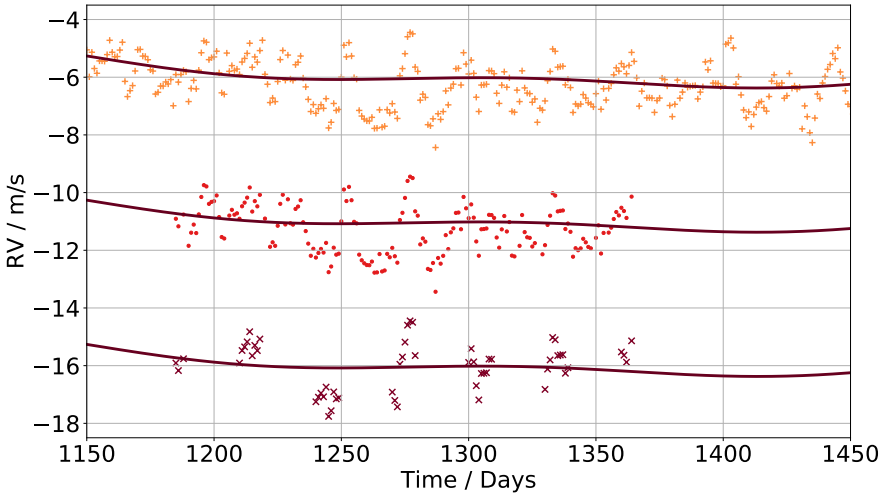


Figure 3.13: A zoom of one of the flatter regions of Fig 3.12 showing the different observation schedules and demonstrating their relative density. In the space (top) and Terra Hunting schedule (middle), a ~25 day period signal can be seen, a result of a large spot group. The reference schedule (bottom) has clearly failed to sample this signal adequately.

3.4.3 Mitigating the SOAP data

In any standard RV planet hunting exercise, astronomers *do not* simply take the computed RVs and try to fit a Keplerian model. They would fall foul of false positives arising from both the schedule cadence and the quasi-periodic noise in present in the data. As I will discuss in Section 3.5, there are a number of techniques astronomers can use to determine what periodicities are attributed to stellar activity as opposed to planets, and in Section 3.3 I listed a few famous cases where false-positives have been debunked. For this study, I do not make any attempt to mitigate the stellar signals with conventional techniques. I know that some correction and mitigation of stellar signals is possible, but that no correction process is perfect and *will* leave structured residuals in the data. Because this work is not a study of how to best mitigate stellar signals, I have made the assumption that I can approximate the *outcome* of these processes by simply dividing the SOAP RVs by a fixed factor. I justify this bold assumption by reason that this process will leave structure in our data at periods and amplitudes close to those originally from the SOAP data, but at a reduced amplitude. To represent optimistic levels of correction that could be possible with the most modern of techniques and a 10–year RV survey, I use a reduction of the stellar signals by a factor of four. It should be noted that the raw SOAP RV peak to peak is approximately $6\,\mathrm{m\,s^{-1}}$ and is therefore reduced to a peak to peak of $1.5\,\mathrm{m\,s^{-1}}$ and as such is still many times larger than the RV of the Earth–twin and higher than our instrument photon noise.

The value of the reduction factor was decided during the initial run tests of PolyChord on a test case of System 1 as observed by the Terra Hunting schedule. Various values of the SOAP reduction factor were tested linearly from 1 through to 10. The parameter estimation was very accurate for a reduction of 6 or more as the RMS of the SOAP data is reduced to just below $1\,\mathrm{m\,s^{-1}}$, and the underlying signal can essentially be 'brute-forced' by the sheer number of data points and a Keplerian model.

However, a reduction of a factor of 6 is not feasible. I wanted to chose a value that was optimistic, i.e. predicting a reduction that could be possible in the near future, but was not out of reach of current techniques. Lanza et al. (2018) observed a selection of quiet stars over a two-year period and used a combination of indicators to reduce the standard-deviation of the RV time-series by a factor of two for half of the target list. More recently, Lanza et al. (2019) combined this technique with observations of the magnetic activity of the Sun-as-a-star. The disc-averaged magnetic flux proves to be the best proxy over a timescale of one month and gives a standard deviation reduction factor of 2.8. Whilst these works were completed *after* the work presented in this chapter, it provides a promising argument that the reduction by a factor of four is not unreachable in the near future.

I do not make the claim that this procedure of simply reducing the SOAP data results in an accurate representation of a Terra Hunting Experiment data-set, or is physically justifiable. However I think this is a reasonable process for the purposes of investigating the effect of a quasi-periodic signal dwarfing the planetary signal, and how different observations schedules will respond.

3.4.4 Limitations of using SOAP

Whilst the SOAP data-sets provide a useful tool of generating RVs from spot dominated activity, it has a few limitations. SOAP *only* generates RVs from spots and faculae, and does not include the contribution from p-mode oscillations, granulations, and magnetic cycles. As discussed, I overcome the first two by taking long exposures of 20 minutes to average them out. Magnetic cycles, however, are the key driver of overall activity levels. In my use of SOAP, the average spot number, lifetime, positions, and sizes have a constant average. This would not be a problem if I was only considering a shorter baseline of observations, say a few hundred days, but for 10 years we know the Sun will have nearly completed one full cycle with changing activity throughout.

SOAP is also generally suited to more active stars than one like our ideal candidate. With SOAP, each spot is accompanied by a faculae halo of fixed ratio to the size of the spot, however we know that faculae can exist in isolation also (Shapiro et al. 2014). For very quiet stars there

could well be zero spots on the surface of the star for periods at a time, but there would likely be other magnetic structures such as faculae.

The lack of magnetic cycles in SOAP was also justification for using a 5-year version of each schedule along with the full 10 year series. The activity of the Sun during a quiet phase is *roughly* constant for a few years, so these subsets of data will reveal if it is possible to detect any of the planets or not in a reasonable estimation of Sun-like activity at quiet times.

Another limitation of SOAP is in its computation requirements. In its current implementation, each RV value takes progressively longer to compute which restricted this study to a 24 hour cadence. Having investigated into the source code of the iterative loops, it would likely require a fairly substantial rewrite to reduce this to from an estimated $O(n \log n)$ to $\sim O(n)$.

3.5 Indicators for Stellar Activity

As discussed in Section 3.4.3, I did not use conventional methods to 'correct' my stellar data, in fact I only ambitiously approximated what the *outcome* of sophisticated stellar noise correction may be. For completeness I describe a few mitigation techniques that are commonly used in RV surveys to reduce the affect of stellar oscillations.

We can use a set of indicators to identify any activity arising from the physical processes mentioned in Sections 3.3.1 - 3.3.4. These indicators are either measured directly from the same spectra used to measure the RVs, or are a result of analysing the CCF.

3.5.1 Spectroscopic Indicators

Spectroscopic indicators are derived directly from the each spectral measurement and can be used to infer the presence of high stellar activity.

The *S*-index is "a dimensionless ratio of the emission in the line cores to that in two nearby continuum band-passes on either side" (Hall 2008). The line-cores mentioned are the Ca_{II} H and K at 396.9 nm and 393.4 nm respectively. It is given by

$$S = \alpha \frac{\Psi_H + \Psi_K}{\Psi_V + \Psi_R}, \tag{3.6}$$

where Ψ_H and Ψ_K are the fluxes at the core of the two lines, and Ψ_V and Ψ_R are the fluxes of the continuum to either side of the lines in the 'violet' and 'red' directions respectfully, and α is a normalisation factor which depends on the exact spectral type of the star measured. Increased stellar activity enhances the chromospheric emission which includes, for some stars, the Ca_{II} H and K lines. Hence, a higher S-index corresponds to higher activity.

The $R^{'}_{HK}$-index introduced by Noyes et al. (1984), is an amendment to the S-index which removes the dependency on the spectral type of the star, and of the instrument used. It is given by

$$R^{'}_{HK} = \frac{\Psi^{'}_{H} + \Psi^{'}_{K}}{\sigma T^{4}_{eff}}, \tag{3.7}$$

where σ is the Stefan-Boltzmann constant, and T_{eff} is the effective temperature of the star. The primes on the fluxes denote that the chromospheric contribution of the reference star has been subtracted.

3.5.2 CCF indicators

The CCF contains more than just the velocity of the star at a given time. Its shape and size can indicate the presence of stellar activity.

The Full Width at Half-Maximum (FWHM) is the width of the CCF at half of its fullest depth. The FWHM will grow as the star rotates faster, something usually attributed to young stars. Young stars also exhibit higher activity levels so the FWHM can be used as a general indicator of the stellar activity. Also, a spot crossing the disk will cause an asymmetry of the CCF as it passes from the blue-shifted to the red-shifted side of the star. Many spots will cause the CCF to have a more chaotic and non-Gaussian shape.

The Bisector Span (BIS) of the CCF is a measure of the asymmetry in general. It was first used in the context of exoplanets by Queloz et al. (2001). It is constructed by taking the middle of the CCF at increasing depths to the bottom. A perfectly symmetrical CCF (i.e. a pure Gaussian) will have a BIS that is just a vertical line. Even a *very* quiet star with just granulation will show a non-zero BIS as granulation has a net blue-shift contribution. However even in this case, the signal from the indicators is often so slight that is it not obvious if the RV is due to a low-mass planet or from stellar activity (Desort et al. 2007). The FWHM and BIS are shown together in Figure 3.14.

3.5.3 The FF' Method

The last indicator I mention is the FF' method of Aigrain et al. (2012). This technique only requires time-series photometry of the target and only has two free parameters, which is much simpler than many stellar modelling disciplines. It consists of computing the product of the photometric flux F with its time derivative F'. Modelling these two time-series provides a basic description of a spot crossing the stellar disc because the RV of a star varies with the *cosine* function of the flux deficit ($\propto$ F), whilst the line-of-sight velocity follows a *sine* function and so

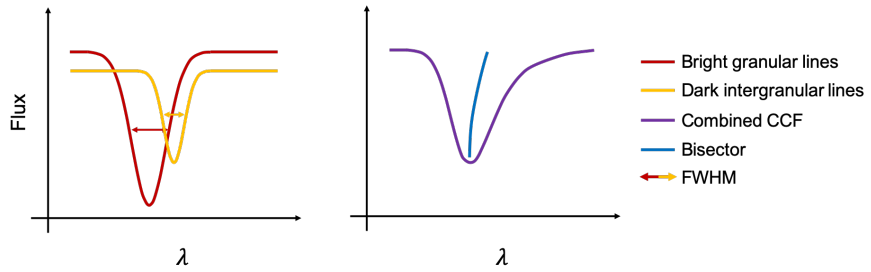


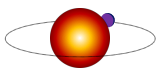
Figure 3.14: The overall shape of the CCF and be used to infer the presence of stellar activity. On the left the two line profiles are from the granulation and the intergranule lanes. On the right the combined CCF profile skew is displayed greatly exaggerated, with the BIS displaying a characteristic 'C' curve due to potential activity. Plot inspired by Haywood (2015).

is $\propto$ F'. Because this technique is simple and fast it can be quickly applied to multiple datasets, however it does not factor in more subtle photometric effects such as the presence of faculae.

3.6 Nested Sampling for RV Analysis

To compare the relative strength of each observation schedule, I needed to extract the planets from the 48 time-series, and compare results. Traditional time-series analysis can be computed with peak-detections on a periodogram from either a Fourier Transform or a Lomb-Scargle analysis. But with the knowledge that the data is noisy, poorly sampled, multi-modal, and contains many superimposed signals, this might not be the best option. Additionally, it would be helpful to know whether the data can be better explained by a certain number of planets over another, and by how much.

I opted to use a nested sampling algorithm that employs Bayesian model selection at its core. Bayesian model selection offers a mathematically rigorous way of calculating the probabilities of models describing data, and the nested sampling algorithm allows the parameters of that model to be efficiently estimated. In the specific case of exoplanets, this means fitting different numbers of planets each with potentially unique parameters, and then statistically weighing up these models for comparison. With this framework it is then possible to *directly* compare the different models and observation schedules. In the case of the Terra Hunting Experiment, this means I can statistically say how well it has performed in discovering a potential Earth-twin when competing against other schedules, and how significant the detections are compared to other models. It is expected that as you increase the quantity and quality of data, the true model will become favoured over the false models. For the rest of this chapter, 'model' refers to the number of planets being fitted to the data. For each of the 48 data sets I fit a 0, 1, 2, 3 and 4 planet model, yielding 240 sets of results in total.



3.6.1 Bayes' Theorem Crash Course

Before I dive into how I use Bayes' theorem with nested sampling, I will provide a brief introduction into why this branch of statistics is useful, and a couple of examples to better understand the use cases of Bayesian statistics.

There are a few ways to express Bayes' theorem but the most common is

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \tag{3.8}$$

where $P(A|B)$ is the conditional probability of event A given event B has occurred, $P(B|A)$ is the conditional probability of event B given event A, $P(A)$ is the prior probability of event A, and $P(B)$ is the marginal probability of event B.

To put this into a simple example[f], lets say that there is a 40% chance it will rain on a Sunday, and if it does there is also a 10% chance it will rain on the following Monday. But if it does not rain on Sunday there is an 80% chance it will rain on Monday. What a weird climate this place has. With frequentest statistics we can calculate the probability it will rain on any given Monday by simply adding up the relevant probabilities of "Raining on Sunday and raining on Monday" and "Not raining on Sunday and raining on Monday" to give

$$(0.4 * 0.1) + (0.6 * 0.8) = 0.52 \equiv 52\%. \tag{3.9}$$

These values are the probability that it rained on Sunday (0.4) *and* the probability that it rained on the following Monday (0.1), plus the probability that it did not rain on Sunday ($1 - 0.4 = 0.6$) *and* the probability that it will still rain on the following Monday (0.8).

Now, what if I asked you the question "It rained last Monday, what is the probability that it also rained the previous Sunday?". Here we can invoke Bayes theorem to help us. We start by writing our probabilities to correspond with the terms for Equation 3.8 such that the probability of it raining on Sunday is $P(A) = 0.4$, the probability of it raining on a Monday we just calculated to be $P(B) = 0.52$, and the probability of it raining on a Monday given that it rained on a Sunday $P(B|A) = 0.1$. We can then see that the probability of it raining on a Sunday *given* that it rained on a Monday is

$$P(A|B) = \frac{0.1 * 0.4}{0.52} = 0.0769 \equiv 7.69\%. \tag{3.10}$$

Bayes' theorem can also be used to assess the accuracy of tests in the face of false positives and false negatives. A classic example is for medical tests with a known rate of false reports[g].

[f] Example values borrowed from the 'simple Wikipedia' entry on Bayes' theorem found here: https://simple.wikipedia.org/wiki/Bayes%27_theorem.

[g] See more information on this example here https://betterexplained.com/articles/an-intuitive-and-short-explanation-of-bayes-theorem/.

For the sake of sensitivity, lets make up a hypothetical disease called PhDitis. From hypothetical historical records, PhDitis can be found in naturally 1% of PhD students around the world, and the best test for it is 80% successful in a diagnosis. However, this test also returns a positive result 9.6% of the time when the disease is *not* present. To answer the question "You have been diagnosed with PhDitis, what are the chances you actually have it?", we can again use Bayes' theorem. A table of probabilities is shown in Table 3.3.

Table 3.3: The test results for either having or not having the hypothetical disease. Bayes' theorem can use these values to help understand the true chances of actually having the disease given a positive diagnosis.

	PhDitis (1%)	No PhDitis (99%)
Positive Result	80%	9.6%
Negative Result	20%	90.4%

Since you have a positive diagnosis, you are somewhere in the top row. The chance that you have the disease (a true positive) is $1\% * 80\% = 0.8\%$, and the chance that you do not have the disease (a false positive) is $99\% * 9.6\% = 9.5\%$. Using Equation 3.8, we can see that the probability of having the disease *given* a positive test can be expressed as

$$P(A|B) = \frac{0.01 * 0.8}{0.095 + 0.008} = 0.076 \equiv 7.6\%. \tag{3.11}$$

The result might be surprising given that our intuition says there should be an 80% success rate on the diagnosis, but we have updated our expectation with the relatively rare occurrence rate of this disease. Simply put, the rate of false positives far exceeds the rate of true positives.

A third, and last, example can be demonstrated with betting. Suppose you are at a Formula 1 race and you want to place a bet between the two championship contenders, Lewis Hamilton and Sebastian Vettel.

Sebastian has won 7 of the previous 12 races whilst Lewis has only won 5. With this information it is clear that a gambling man should place his bet on Sebastian (58.3% victory rate over 41.7%). However, with more information you might be able to place a smarter bet. Suppose the weather forecast predicts that it will rain for the next race, as it did for 4 of the previous 12 races. Of those 4, Lewis won 3 whilst Sebastian only won 1. Our intuition might now favour Lewis as a wet-weather champion but how much should we trust it?

Again with Equation 3.8, we can write the terms as: the probability that Lewis wins in the rain is $P(A|B)$, the probability that it is wet when Lewis has won $P(B|A) = 3/5 = 0.6$, the probability that Lewis will win any race $P(A) = 5/12 = 0.417$, and the probability that those events took place during rain $P(B) = 4/12 = 0.333$. As before, we then compute the probability

of Lewis winning given that it is raining as $P(A|B) = \frac{0.6*0.417}{0.333} = 0.75$. Lewis's chances of victory have gone up, and probably sufficiently to now want to back him in the upcoming race.

Bayes' theorem allows us to change our prior beliefs with some evidence, and update them to become our posterior beliefs. In the context of fitting a Keplerian signal to RV data, we can calculate the likelihood that a N-planet model fits the data *given* some assumption (e.g. a noise model or a least squares regression), and look at the most probable values of the parameters of that model.

3.6.2 Using Nested Sampling for RV Analysis

Whilst making educated bets on the horses or predicting the outcome of races is an interesting application of Bayes' Theorem, it does not help with deciding which astrophysical model best describes some stellar radial velocity data.

The formulation of Bayes' Theorem of Equation 3.8 can be re-written in a more useful form that allows us to calculate the probability of mathematical models with parameters, instead of just pure probabilities of various numbers. We can then compare how likely different models are at describing some data, *and* estimate the values of the parameters of that model. In the case of trying to find a particular Keplerian signal in some radial velocity model, we can use Bayes' Theorem to tell us how many planets are the most likely (the model), and what each of their periods, amplitudes, and phases are (the parameters of the model).

I follow the notation and methodology of Feroz et al. (2011). Given data D and a model H with parameters Θ, Bayes' theorem states:

$$\mathrm{P}(\Theta|D,H) = \frac{\mathrm{P}(D|\Theta,H)\mathrm{P}(\Theta,H)}{\mathrm{P}(D|H)} \;\leftrightarrow\; \mathcal{P}(\Theta) = \frac{\mathcal{L}(\Theta)\pi(\Theta)}{\mathcal{Z}} \tag{3.12}$$

where $\mathcal{P}(\Theta)$ is termed the posterior distribution of the parameters, $\mathcal{L}(\Theta)$ the likelihood, $\pi(\Theta)$ the prior and $\mathcal{Z}$ the Bayesian evidence. After specifying a model via its likelihood and prior, one may numerically sample the posterior and compute the evidence using nested sampling (Skilling 2006).

In this new notation, the Likelihood ($\mathcal{L}$) is the probability to obtain the data (D), given some model (H), that has some parameters (Θ). It can be expressed mathematically in almost any form as the user desires, but a typical description takes the form of a least-squares minimalisation. In the example of finding a Keplerian signal in a time-series, the Likelihood would describe a normal distribution of values about some sinusoidal model with parameter values (Θ). The Prior (π) is our knowledge of the parameters (Θ) of the model (H) *before* we take into account the data. Again, we mathematically describe each parameter's prior distribution as per the user requirements. Typically these take the form of uniform or normal distributions between some

Table 3.4: The Jeffreys scale for interpretation of model probabilities, Jeffreys (1983).

$\lvert\Delta \ln R\rvert$	Odds	Probability	Remarks
< 1.0	$\lesssim$ 3:1	<0.750	Inconclusive
1.0	~3:1	0.750	Weak Evidence
2.5	~12:1	0.923	Moderate Evidence
5	~150:1	0.993	Strong Evidence

upper and lower bound. The Bayesian evidence $\mathcal{Z}$ is a normalisation factor that ensures the posterior probability lies between 0 and 1.

The posterior distribution is used to perform *parameter estimation*, namely the quantification by $\mathcal{P}(\Theta)$ of our knowledge of a model's parameters in light of the data and our prior assumptions.

The evidence $\mathcal{Z}$ is used to perform *model comparison*. Applying Bayes' theorem again to a sequence of models $\{H_0, H_1, \ldots\}$:

$$\mathrm{P}(H_i|D) = \frac{\mathrm{P}(D|H_i)P(H_i)}{P(D)} = \frac{\mathcal{Z}_i \phi_i}{\sum_k \mathcal{Z}_k \phi_k} \tag{3.13}$$

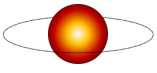
we can see that the evidences $\mathcal{Z}_i = \mathrm{P}(D|H_i)$ along with the model priors $\phi_i = \mathrm{P}(H_i)$ can be used to infer the relative likelihood of a model H_i within the set of models. The model priors are typically taken to be uniform, and one can compare the relative probabilities of two models via a Bayes factor:

$$R = \frac{\mathrm{P}(H_1|D)}{\mathrm{P}(H_0|D)} = \frac{\mathrm{P}(D|H_1)\mathrm{P}(H_1)}{\mathrm{P}(D|H_0)\mathrm{P}(H_0)} = \frac{\mathcal{Z}_1 \phi_1}{\mathcal{Z}_0 \phi_0}. \tag{3.14}$$

The natural logarithm of R provides us with a convenient measure of what constitutes a significant difference between two models and is summarised in Table 3.4. If, for example, you were fitting Keplerian models to some RV data and your 2-planet model versus a 3-planet has a $\Delta \ln R = 5.1$, you would take that as strong evidence that the two planet model is a better description of the data.

Nested sampling provides an efficient tool for parameter estimation especially in the context of Bayesian statistics (Skilling 2006). Computing the evidence, $\mathcal{Z}$, is very computationally intensive set of integrals often over a wide parameter space with high dimensionality. To efficiently sample this posterior and calculate the evidence at each sample, nested sampling is used. Generally speaking, nested sampling is a technique where the evidence is efficiently calculated through a Monte Carlo approach, and the posterior distribution is a produced as a natural by-product. For testing models against data, nested sampling not only provides a measure of model selection, but the efficient calculation of the parameters of that model along with errors.

Nested sampling, then, allows us to statistically estimate the multidimensional posterior landscape without explicitly calculating it. This saves a vast amount of computation time which



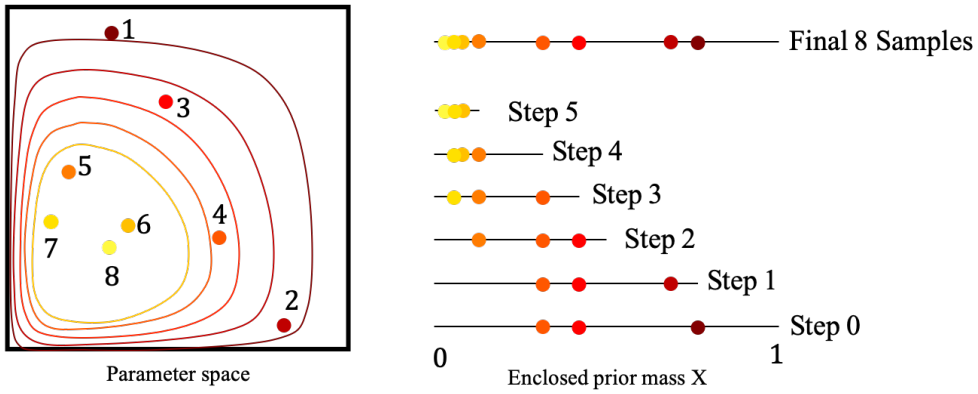


Figure 3.15: An example of a 2-dimensional posterior space being approximated by a series of points that are procedurally generated with increasing likelihood (red to yellow). At each step the worst point (lowest $\mathcal{L}$) is removed and a new one is generated. The information stored by each point allows the posterior distribution to be sampled and its peak efficiently estimated.

becomes increasingly important when the dimensionality increases to high numbers (e.g. a 4 planet Keplerian model described by 6 parameters per planet is a 24-dimension problem!). Skilling (2006) first described the nested sampling method as an efficient tool of parameter estimation.

Suppose we have a 2-dimensional likelihood distribution and we are attempting to find the most likely solution - the peak. A simple nested sampling algorithm might generate 3 points from the prior with which to evaluate $\mathcal{L}$. The worst point (lowest $\mathcal{L}$) is removed, and a new point is generated with the criteria that it has a higher likelihood than the previously dropped data point. This process is repeated until all three points agree on the approximate maximum likelihood with some pre-defined precision limit. At each step, the new point's location can be randomly generated or drawn from a distribution. How exactly to draw these points depends on the application, the number of points, the dimensionality etc, and is not part of the scope of this research. This is graphically depicted in Figure 3.15.

With nested sampling, an approximation of the posterior can be efficiently calculated in sensible time-frames. Then, the posteriors of a particular model are used to compute model comparison in the context of Bayes' Theorem.

To perform nested sampling, I used the implementation provided by PolyChord (Handley et al. 2015a,b). The nested sampling algorithm MultiNest (Feroz & Hobson 2008; Feroz et al. 2009, 2013) has also been successfully applied to both real and synthetic exoplanet RV data (Feroz et al. 2011), and was also used in the recent RV Challenge (Dumusque 2016; Dumusque et al. 2017). PolyChord is a successor to MultiNest, designed to work more efficiently with a higher number of dimensions, a feature that becomes increasingly important for a multi-planet system with many parameters per planet.

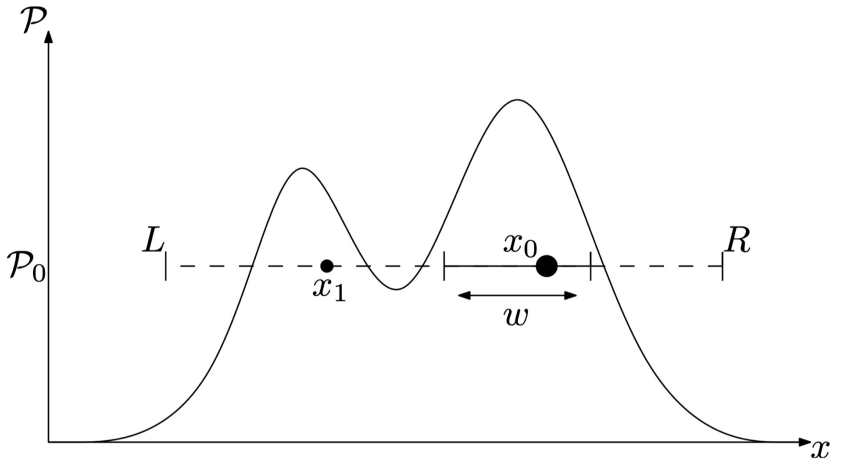


Figure 3.16: The slice-sampling method employed by PolyChord. The bounds L, R are set on the first slice. PolyChord samples the posterior with multiple slices of random directions and starting points, and can be ran in a parallelised mode to utilise the architecture of modern CPUs, figure from Handley et al. (2015a)

PolyChord samples the posterior with a technique called 'slice sampling'. Here, a one-dimensional cross section of posterior distribution is considered, and a horizontal line is drawn across. From an initial point, x_0, within the slice a new point, x_1 is generated with a distribution of $P(x_1|x_0)$. This is depicted in Figure 3.16.

Nested sampling outperforms typical MCMC methods in situations where the sampling space is large, multidimensional, and non Gaussian, which often causes MCMCs to find local, rather than global, maximum solutions. To operate in these scenarios, MCMCs often require precise fine-tuning of their parameters (step size, direction, distribution etc), and are computationally inefficient when denser point distributions are required.

3.6.3 MultiNest versus PolyChord

As mentioned, MultiNest has been successfully used in Exoplanetary science whilst the more recent PolyChord has so far only been used in cosmological disciplines such as analysing the Planck datasets.

However, PolyChord was chosen over MultiNest for reasons of scalability and future-proofing with reliable updates and fixes in the foreseeable future. Whilst MultiNest is faster than PolyChord for sampling up to approximately 64 dimensional spherical Gaussian posteriors, this crossover threshold is significantly lowered for more complicated posterior shapes.

For my exoplanet distributions, the crossover was found to be between 10 and 20 dimensions. Preliminary runtime tests of MultiNest were between 2 and 3 times longer on average. In these tests I also examined the posteriors of simple circular orbit parameters of some test–case planets.

Table 3.5: The runtime tests between MultiNest and PolyChord for some circular planets sampled with a Terra Hunting Schedule. There are 3 parameters per planet. PolyChord clearly scales better with increasing dimensionality.

Number of Planets	MultiNest Time / s	PolyChord Time /s
1	7.03	12.44
2	127.36	91.86
3	747.45	458.73
4	4556.20	1396.72

I found that as the number of planets were increased, MultiNest results started to deviate from the true values, but were recovered correctly by PolyChord. Moreover, MultiNest requires a relatively high minimal number of live points n_{live}~400, whilst PolyChord produces accurate posteriors with far lower numbers n_{live}~$5n_{\text{dims}}$. Lastly, I found that MultiNest run-times scale very poorly with increasing N_{dims}, Table 3.5 shows some of the runtime timing results. MultiNest was faster than PolyChord for *only* a one planet model, and also produced poorer posteriors (broad range, multi-modal, non Gaussian) for all models, Figure 3.17. Running the analysis of all observation schedules applied to *one* solar system using PolyChord took at most 57 hours on a single core of a 2.90GHz Intel Xeon E5–2690. There is capacity to speed this up by a factor of $10 - 100$ for future pipelines by re-writing the likelihoods in C/C++ rather than Python. This implementation of PolyChord was not parallelised, but since the analysis between schedules and solar systems is independent, the analysis scripts were manually sent to different cores on the Cavendish Astrophysics Cluster to massively speed up the process.

3.6.4 Defining the Likelihood and the Priors

With the datasets generated as per the procedure of List 3.2, and a model ready to test from Equation 3.3, only two more things need to be defined before the nested sampling can commence: the Likelihood and the Priors.

In my first iteration of the analysis, I chose to omit the RVs generated by SOAP. This deliberately unrealistic data-set acted as a test of PolyChord in the simplest case: can we identify a poorly sampled sinusoidal signal with some Gaussian noise added on top?

In this paradigm, we can formulate our likelihood function as a minimisation of a least-squares fit as

$$\ln \mathcal{L}(\Theta) = \sum_i -\ln\sqrt{2\pi\sigma_i^2} - \frac{1}{2\sigma_i^2}[v(t_i;\Theta) - v_i]^2, \tag{3.15}$$

where v_i and σ_i are the ith RV measurement and the associated error, $v(t_i;\Theta)$ is the ith RV

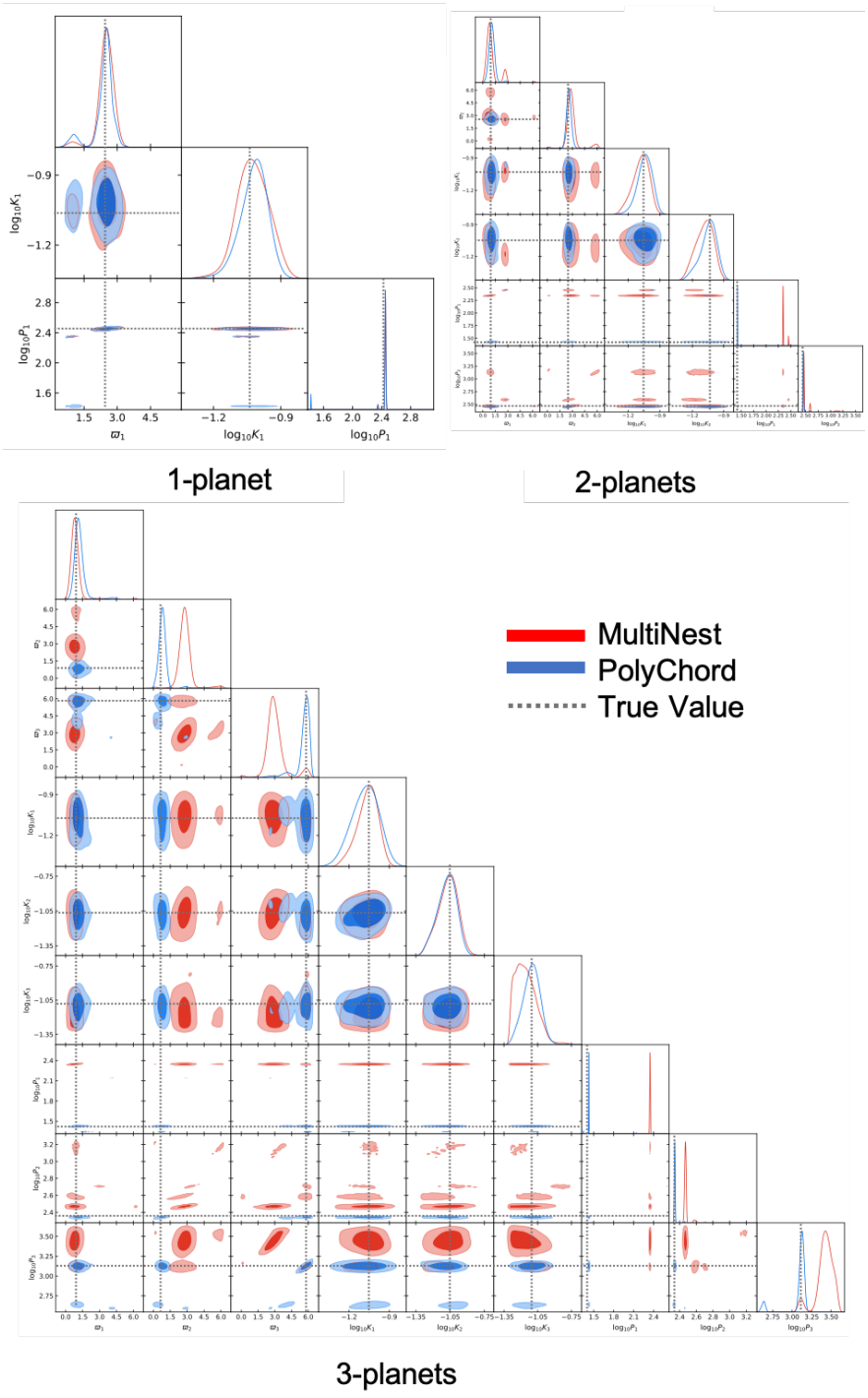


Figure 3.17: Posteriors from MultiNest and PolyChord of a 1, 2 and 3 planet model. For higher dimensionality, PolyChord performs much better and produces more accurate parameter estimates. Here, 1, 2, and 3 circular planets with low white-noise were sampled with a Terra Hunting Experiment schedule. The parameters ω_i, K_i, and P_i are the phase, RV Semi Amplitude and Period respectively for the i^{th} planet.

predicted from a model (Equation 3.3) with parameters Θ, (MacKay 2003). With the stellar signals omitted, there is only a normal distribution of noise around the underlying signal of the planets hence our likelihood only needs minimise the least squares between the model and the data.

I then include the stellar RVs from SOAP and repeat the analysis. This will let me investigate how the variability has affected the rates of false positives across the different schedules and solar systems. For this part, I leave the model and likelihood function the same. Hence, I am still telling PolyChord that the data can still be described by a Keplerian signal only, and that the errors are purely Gaussian. Whilst this statement is incorrect, the stellar signals have already been 'corrected' (Section 3.4.3), and so I am assuming that any remaining periodicity is either an artefact of the correction process or *is* a Keplerian signal.

The priors, our prior belief of the range of the parameters of our model, are defined by the user. PolyChord allows for a variety of prior distributions from which to sample, and at each stage of the nested sampling it will draw a value from this distribution to test against the data with the Likelihood function. A prior distribution can be as generic or specific as needed, however a more specific distribution could introduce bias into the sampling. For example in the case of Exoplanet discovery, the prior on the RV semi amplitude would not need to weight values up at many $\mathrm{km\,s^{-1}}$ as high as those at a few $\mathrm{m\,s^{-1}}$ as the former signals are rarer due to a lack of planets at those high masses. In this example, a user could chose to define their RV semi amplitude prior as a normal distribution, or a log-normal distribution to weight these extreme values accordingly. The exact function used should be physically justifiable and should represent the true distribution of known planets as accurately as possible. A user searching for exoplanets could look at all of the known exoplanets and fit an envelope for various parameters such as mass, period, eccentricity etc and use those as a 'sensible' prior. However, because our current demographics of all known exoplanets include some severe observational biases (instrument thresholds, targeted surveys etc) these priors would be inherently biased. There are alternative prior distributions that a user could choose to use that offer more functionality than a simple uniform distribution, but contain minimal biases when assuming the true distribution of the posterior. In Gregory (2007b) the authors discuss the choice of priors with consideration of the physics of orbital mechanics. The lower bound of period need not be less than 1 day due to Roche Lobe considerations (the planets orbit starts to decay to zero beyond here), and the upper bound can be less than 1000 years because at these timescales nearby stellar neighbours may start to affect the velocity. The Jeffreys prior is a sensible distribution when dealing with many orders of magnitude as it has a uniform probability distribution in frequency (1/period). In Feroz et al. (2011) the authors used a *modified* Jeffreys prior which has a near uniform distribution for the parameter below some value, but acts like a log-uniform above some value.

Table 3.6: Specification of prior probability distributions on the parameters.

Parameter	Prior	Lower Bound	Upper Bound
P (days)	Log Sorted Uniform	10	4000
K ($\mathrm{m\,s^{-1}}$)	Log Uniform	0.05	2000
ϖ (radians)	Uniform	0	2π

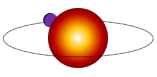
It has the mathematical form of

$$Pr(\theta|H) = \frac{1}{(\theta + \theta_0)\ln(1 + \frac{\theta_{max}}{\theta_0})}, \tag{3.16}$$

where θ is the value of the parameter, θ_0 is the threshold value between uniform and log-uniform, and θ_{max} is the upper bound of the prior distribution. Despite the relative success of this prior in Feroz et al. (2011) and others, it still imposes physical constraints on the priors. Also, should I use this prior for my purposes, I would still likely set my θ_0 to be around 10 years such that I am log-uniformly sampling for periods across the entire range of my 10 year data sets.

For these reasons, I elected to keep my priors as unbiased as possible. For the phase of circular orbit of the planet is is simply uniform between 0 to 2π radians. For the RV semi amplitude, the prior is uniform in log space between $0.05\,\mathrm{m\,s^{-1}}$ to $2000\,\mathrm{m\,s^{-1}}$. Having the RV semi amplitude in log-space puts some preference to lower amplitude planets (something I knew a priori was true for my data) whilst still allowing for the fitting of high mass planets. The period prior is also uniform in log-space but I place a further constraint of sorting it. This choice was to break the switching degeneracy between between sets of planet parameters (Gregory 2007a), we choose to impose a prior constraint on the periods such that first planet has the smallest period, followed by the second and so on. This is a physically motivated prior as we know that planets can not share orbits, so there is no need to check a particular period value twice *if* a planet has already been found there. This ordered prior constraint could still be vulnerable to signals of very similar periods and amplitudes, e.g. two signals of 10.00 days at $1\mathrm{m\,s^{-1}}$ and 9.99 days at $1\mathrm{m\,s^{-1}}$ could be fitted by a single curve of 10 days at $2\mathrm{m\,s^{-1}}$, but it reduces the degeneracy from $N!$ to just N.

This initial test serves a few purposes. It allows for first tests of PolyChord with a simple model, likelihood, and datasets. Plus the results will be an early indicator of the performance of PolyChord *and* of the various observation schedules when the data is as simple as possible.



3.6.5 Using PolyChord

To actually use PolyChord I setup my code in script batches with Python. Each script had access to all of the simulated RV data from the various solar systems and schedules and would load a pre-determined subset for analysis. PolyChord would then be called to test the Keplerian model on those datasets and record the results. Within a script I would loop the model to search for $N_p = 0, 1, 2, 3,$ and 4 in turn. PolyChord would then conduct the nested sampling, generate posterior distributions and model evidences for later analysis where I use separate scripts to plot the results. PolyChord is distributed with some very easy to use Python functions that call the underlying fortran code. An example of the simple use of PolyChord with Python is below. Here, I only need to define a function which is the model I want to fit to the data, my likelihood function, and my priors. Given some set of time-stamps t and some data y, PolyChord will commence the nested sampling with just a few more lines of code which are omitted here[h].

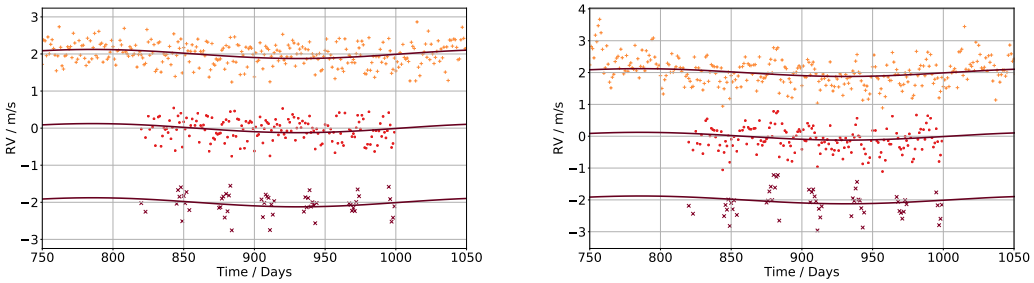
```
# model of a single sinewave with 100 datapoints
t = np.linspace(0,100,num=100)
def f(t, a1,w1):
    return a1 * np.sin(w1 * t)

# least-squares minimisation - only Gaussian noise present
def likelihood(theta):
    a1, w1 = theta
    logl = -log(2*pi*sig**2) *
           ndat/2. -sum((y - f(t,a1,w1))**2)/ sig/2.
    return logl, []

# simple uniform priors
def prior(x):
    a1 = UniformPrior(0,2)(x[0])
    w1 = SortedUniformPrior(0,5)(np.array(x[1]))
    return [a1,w1]
```

In the above example, `ndat` is the number of data points in total for the RV array of length `len(y)`, and `sig` is the σ_i measurement error on each data point. The two prior functions are imported from a priors file where user can use a predetermined prior or write their own. In

[h] These include a few hundred lines of reading and writing data, setting path-names, plot parameters, PolyChord output settings etc



(a) Without SOAP (b) With SOAP

Figure 3.18: System 1 observed by the three schedules with and without the SOAP data. The space schedule is top, Terra Hunting is middle, and the reference schedule is at the bottom, in all cases the underlying Keplerian signal is the solid line. In the case where the SOAP data is present, it has been corrected by a fixed value of 75% as discussed in Section 3.4.3.

the case of a Log-sorted-uniform prior for the RV Period prior, I wrote my own with the help of the PolyChord author W. Handley, my two other prior functions were already available.

3.7 First Results - Only Gaussian Noise

In this section I discuss the results of applying PolyChord to all of the schedules observing the 4 solar systems but omitting the SOAP data and allowing PolyChord to search all planetary models unconstrained. As mentioned, this acts as a simple test of the analysis pipeline itself, plus it will act as a benchmark to compare the results when the full data-set is analysed. These tests were carried out to assess the absolute best-case scenario of perfect mitigation of the stellar noise, and therefore are purely a test of the observation schedules only. I will go through each solar system in term and discuss the model selection results first, then look at the favoured parameter estimates from each schedule.

3.7.1 System 1 Results - Gaussian RVs, Unconstrained Model Search

Solar system 1 only contains a lone Earth-twin on a 293 day orbit. For all schedules the 1-planet model is strongly favoured. However, it is not just the identification of the correct model that matters, the posterior distributions of the parameters also need to agree with the injected values for a 'detection' to be confirmed. Both sets of data (with and without the SOAP RVs) are shown in Figure 3.18.

In terms of model comparison, the reference schedules both favoured the true $N_p = 1$ model. For the 5-year variant the next favoured model was the $N_p = 2$ model with a $|\Delta \ln R| = 3.14 \pm 0.20$ which constitutes 'moderate to strong evidence' that the $N_p = 1$ model the best description

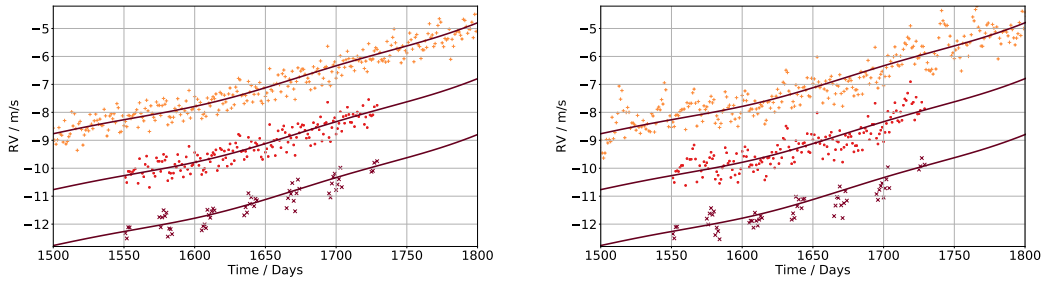
of the data. For the 10-year variant the $N_p = 2$ model again was the next favoured with $|\Delta \ln R| = 3.93 \pm 0.23$ which is approaching 'strong evidence' that the $N_p = 1$ model best describes the data.

For the reference schedule's parameters, all of the parameter posteriors contain the true value within a 1σ error. However the errors on the period are very large: 292 ± 82 d and 272 ± 70 d for the 5 and 10 year schedule respectively. This result is unsurprising when considering the Gaussian noise is ~3 times that of the signal ($1\sigma = 30\,\mathrm{cm\,s^{-1}}$), and the data is very poorly sampled. Even without any stellar activity at all, the reference schedule has a poor constraint on the period of the planet. For the five year schedule, the RV semi amplitude is underestimated at $10 \pm 3\,\mathrm{cm\,s^{-1}}$ but the 10-year schedule returns a good estimate at $11 \pm 2\,\mathrm{cm\,s^{-1}}$. For the phase of the orbit, the 5-year schedule returned a more accurate and precise value as compared to the 10-year schedule 3.23 ± 0.47 rad and 3.08 ± 0.78 rad respectively. This result could just be a coincidence that the particular cadence of the 10-year schedule favoured a slightly out-of-phase curve. In the simplest case of a single planet and no stellar signals to confuse the analysis, the reference schedule already shows signs of struggling to detect the Earth-twin signal.

The $N_p = 1$ model is favoured by both of the Terra Hunting schedules. For the 5-year schedule the next favoured is the two planet model with a $|\Delta \ln R| = 3.32 \pm 0.24$, and similarly the for 10-year schedule the two planet was the next nearest with $|\Delta \ln R| = 5.53 \pm 0.24$. This is a significant improvement over the reference schedule.

For the Terra Hunting schedule's parameters, the posteriors are much narrower and more accurate than those of the reference schedule. For the 5-year variant, all of the posteriors agree to the true value within error apart from the orbital phase which was underestimated. The RV semi amplitude and period had good estimates of $11 \pm 1\,\mathrm{cm\,s^{-1}}$ and 286.47 ± 7.45 d respectively whilst the phase was calculated as 2.88 ± 0.28 rad which is an underestimate of over 2σ. For the 10-year variant the estimates improve to K=$10 \pm 1\,\mathrm{cm\,s^{-1}}$, P=$290.89 \pm 1.25$ d, and ϕ =3.13 ± 0.19 rad. Here is early confidence that the Terra Hunting schedule is a clear improvement over the reference schedule as with 3 times the data, the estimates have improved and the error on the period has massively decreased by more than a factor of $\sqrt{3}$. However, the slight under-fitting of the phase is a problem common with almost all of the following low-mass planets I discuss from here on, even with the 10-year space schedule mentioned later. Whilst the phase values are 'close', this difference equates to a timing error of many tens of days if it were used to predict the occurrence of a transit for follow-up work. However, in the simplified Keplerian regime, a circular orbit is quite insensitive to even a moderate phase offset. Whereas an eccentric curve has an asymmetrical profile, hence a more realistic model with this parameter would likely yield a better fit for the phase, see Figure 1.2.

Lastly, for the space schedule, the 5-year case strongly favoured the $N_p = 1$ model with



(a) Without SOAP (b) With SOAP

Figure 3.19: As per Figure 3.18 but with System 2 observed by the three schedules with and without the SOAP data. The 300-day window displayed is chosen to show the slight wobble of the Earth-twin planets in the presence of a large gas giant.

$|\Delta \ln R| = 5.68 \pm 0.26$ and the 10-year case even more so with $|\Delta \ln R| = 8.52 \pm 0.29$. This is not a surprising result given the complete coverage across both schedules. For the first time I start to see some highly significant evidence rejection with certain models, for example the 10-year space schedule rejects the $N_p = 0$ model with a $|\Delta \ln R| = 109$.

The posteriors have again improved from the previous schedule. For the 5-year variant, $K = 11 \pm 1\,\mathrm{cm\,s^{-1}}$, $P = 290.17 \pm 2.16\,\mathrm{d}$, and $\phi = 3.28 \pm 0.17\,\mathrm{rad}$. These results are comparable in accuracy and precision to the 10-year Terra Hunting schedule which could indicate the benefit of a shorter and denser schedule over a longer and more sparse schedule where both have roughly the same number of data points (see Table 3.2). However this conclusion could change with the inclusion of the SOAP data. For the 10-year variant $K = 11 \pm 1\,\mathrm{cm\,s^{-1}}$, $P = 292.10 \pm 0.08\,\mathrm{d}$, and $\phi = 3.41 \pm 0.12\,\mathrm{rad}$. These values find the true parameters with good accuracy and high precision. Only the period estimate 'misses' the true value of $P = 293$ days due to a tiny fractional error of ~0.03%.

The full tabulated results of system 1 can be found in the Appendix Table A.1.

3.7.2 System 2 Results - Gaussian RVs, Unconstrained Model Search

System 2 is an analogue of our own Solar System but with just 3 planets contributing to the Keplerian signal, a Venus, Earth, and Jupiter analogue. Here, even with just Gaussian noise, is is expected that this combination of signals will provide a source of false positives for the more sparse schedules. The RVs for system 2 are shown in Figure 3.19a.

The reference schedules both favoured an incorrect model of $N_p = 2$ and $N_p = 4$ for the 5 and 10-year schedule respectively. For both schedules, the $N_p = 1$ and $N_p = 3$ models were also of near equal significance indicating that PolyChord could fit a variety of planet combinations with similar evidence.

Both reference schedules found the Jupiter-analogue with moderate errors but both also failed to confidently find either of the terrestrial companions. There is evidence of a Venus and Earth-analogue detection for the 10-year variant with periods of $P = 159.37 \pm 56.51$ d and $P = 341.13 \pm 378.18$ d respectively. But the very large errors place low confidence on these signals. The 10-year variant finds a 4$^{\text{th}}$ planet at $\sim 1/2$ of the period of the Jupiter-analogue but at low amplitude, an indication that it is trying to fit this false positive to the periodicity found in the aliases of the observation schedule.

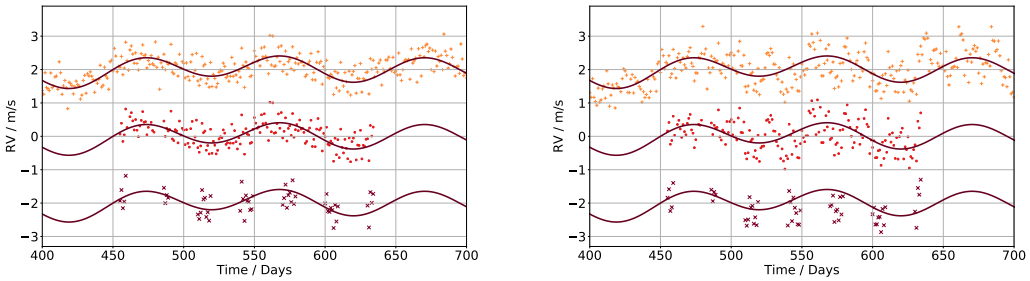
The Terra Hunting schedules performed better and favoured the true $N_p = 3$ planet model. However, in both cases, the $N_p = 4$ model is within $|\Delta \ln R| = 3.82 \pm 0.46$ and $|\Delta \ln R| = 1.49 \pm 0.94$ for the 5 and 10 year schedule respectively. The fact that the 10-year schedule favoured the true model to *less* significance to a false model than the 5-year schedule is interesting. It is likely that aliases of the observation schedule window function, or of existing planets are providing strong evidence in false positives. Also for the 10-year schedule, all other models were more *disfavoured* than in the 5-year variant which is a positive result.

Both find good estimates for the Jupiter-analogue, even the 5-year schedule which has $\sim 1/2$ of its period sampled. Both also find the Earth-twin with the 10-year schedule placing it at $K = 12 \pm 1\,\mathrm{cm\,s^{-1}}$, $P = 294.78 \pm 1.96$ d, and $\phi = 4.77 \pm 0.20$ rad - all within 2σ of the true value. The 5-year Terra Hunting schedule found a false positive planet at $P = 767.69 \pm 686.73$ d with an RV semi amplitude of $K = 0.20 \pm 0.38\,\mathrm{cm\,s^{-1}}$, both of these large errors indicate that the solution is very poorly constrained and is likely being fitted to aliases of the observation schedule.

The space schedules performed well also. They both favoured the true $N_p = 3$ model with very high significance over lower number of planets, but like the Terra Hunting schedules were fairly close to the $N_p = 4$ planet models. Again, I attribute this incorrect model selection to aliases within the data, it is likely that a false-positive planet at a period of an integer fraction of one of the true planets could give strong evidence along with the 3 true planets.

Both of the 5 and 10-year schedules found good estimates of the true periods of the 3 planets and all lay within 2σ of the true values, *and* the distributions were very narrow. As with system 1, the 5-year space schedule found values very similar to that of the 10-year Terra Hunting schedule. This is further indication that a shorter but denser schedule could compete with a longer and sparser schedule but the truth of this statement is obfuscated with the lack of stellar signals from SOAP2.0.

The full tabulated results of system 2 can be found in the Appendix Table A.2.



(a) Without SOAP

(b) With SOAP

Figure 3.20: As per Figure 3.18 but with System 3 observed by the three schedules with and without the SOAP data. The 300-day window displayed is chosen to show the combination of 3 similar mass planets on different periods.

3.7.3 System 3 - Gaussian RVs, Unconstrained Model Search

System 3 contains 3 Earth-twins at different periods of 101, 197, and 293 days. As with system 2, it is expected that the similar signals would force false positives for the under sampled schedules. Both sets of data (with and without the SOAP RVs) are shown in Figure 3.19.

Both of the reference schedules favoured the $N_P = 4$ model with low to moderate significance over the true $N_P = 4$ model. In particular, for the 5-year schedule the $N_P = 2$ model was also only marginally unfavoured. Both of the reference schedules managed to find the large Jupiter-analogue and struggled with the lower amplitude signals. In this system, the only real change is the effective swap of the Jupiter-analogue for an Earth-twin at 101 days, so it is not surprising that they have struggled again here.

The 5-year reference schedule only successfully found the inner most planet with moderate success ($K = 16 \pm 4\,\mathrm{cm\,s^{-1}}$, $P = 99.08 \pm 9.87\,\mathrm{d}$, $\phi = 3.30 \pm 0.46\,\mathrm{rad}$). The other three planets are false positives all at a similar phase and with very large errors, over 50%, indicating that PolyChord could not decide where in phase-space to place these signals as their amplitude is very low. This particular schedule has a repeating observation pattern of 10 observations every 30 days for 180 days, and therefore aliases at integer multiples of these periods. This could explain the values of the rest of the periods found ($P = 167.30 \pm 38.65\,\mathrm{d}$, $P = 248.01 \pm 51.82\,\mathrm{d}$, $P = 508.95 \pm 231.93\,\mathrm{d}$) but the errors are so large it could just be fitted to the noise. The 10-year variant interestingly only found the planet at 197 days and missed the others. However the true parameters for the other two planets are within error of 2 of the other 3 candidates favoured by this schedule, but the errors are very large so a 'detection' is not conclusive.

For the Terra Hunting schedules, both favoured the true $N_p = 3$ model with strong significance over all others apart from the 5-year schedule had a $|\Delta \ln R| = 1.21 \pm 0.27$ for $N_p = 4$.

Both the 5 and 10-year schedules found the three Earth-twins with good estimates on

the parameters and low errors. For the 297-day candidate, the 5-year schedule found $K = 11 \pm 2\,\mathrm{cm\,s^{-1}}$, $P = 294.97 \pm 3.76\,\mathrm{d}$, and $\phi = 1.85 \pm 0.27\,\mathrm{rad}$. And the 10-year schedule found $K = 11 \pm 1\,\mathrm{cm\,s^{-1}}$, $P = 292.32 \pm 1.24\,\mathrm{d}$, and $\phi = 1.71 \pm 0.20\,\mathrm{rad}$. For both of these, the RV semi amplitude and period are very good estimates of the parameters and the errors are of order 1%, but the phase is still underestimated for the 10-year schedule. This could be a consequence of the least-squares likelihood function having a tendency to under-fit parameters when over half of the data is missing.

For the space schedules, both favoured the true $N_p = 3$ model with very strong significance, but like the 5-year Terra Hunting schedule the $N_p = 4$ model was within moderate significance at $|\Delta \ln R| = 2.14 \pm 0.38$.

Both the schedules estimates within error for the three planets and improved estimates on the phase as compared to the Terra Hunting schedules. The 10-year schedule found well constrained parameters for the Earth-twin at 297 days with $K = 11 \pm 1\,\mathrm{cm\,s^{-1}}$, $P = 292.46 \pm 0.80\,\mathrm{d}$, and $\phi = 1.78 \pm 0.12\,\mathrm{rad}$. It could be concluded here that the large 6-month gaps in the Terra Hunting schedules leave the phase more unconstrained for a period similar to the observation window size, whilst this does not affect the amplitude and period as much.

The full tabulated results of System 3 can be found in the Appendix Table A.3.

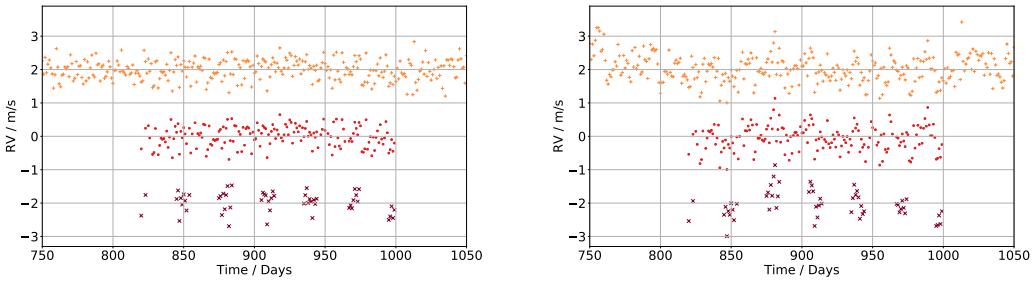
3.7.4 System 4 - Gaussian RVs, Unconstrained Model Search

System 4 contains no planets and is a test of PolyChord in determining the null hypothesis and an examination of which schedules find any false positives that are purely a result of either the observation schedule cadence, or the SOAP RVs.

Interestingly, both the reference schedules were the only ones to favour the true $N_p = 0$ model with very strong significance. It could be that the sheer *lack* of data and the poor sampling meant that no false positives could be found in the sparse, normally distributed data.

The Terra Hunting schedules both favoured $N_p = 1$ with moderate significance, and the $N_p = 0$ true model was not far behind in 2nd and 3rd place for the 10 and 5-year schedule respectively. The false positives found here have large errors and the RV value sits at the lower end of the RV prior which indicates PolyChord is simply trying to fit the model to noise. The periods have large errors $P_{5\mathrm{year}} = 162.80 \pm 107.97\,\mathrm{d}$ and $P_{10\mathrm{year}} = 240.09 \pm 155.15\,\mathrm{d}$ which is further evidence that these unconstrained signals have been forced onto the data.

The space schedules also found the $N_p = 1$ model to be the favourite, this time with strong significance over the others. The 5-year schedule rejected the true model with $|\Delta \ln R| = 35.18$, whilst the 10-year variant rejected it with $|\Delta \ln R| = 9.89$. As with the Terra Hunting schedules, the single false positive was at the lower end of the RV prior, and had large unconstrained errors for both period and phase. It could be that this sheer amount of data allows for some signal to



(a) Without SOAP (b) With SOAP

Figure 3.21: As per Figure 3.18 but with System 4 observed by the three schedules with and without the SOAP data. There are no planets in this data so no underlying model is shown.

be brute-force fitted to the Gaussian noise, but the large errors of the period and phase indicate that the solution is likely meaningless.

The full tabulated results can be found in the Appendix Table A.4.

3.7.5 Conclusions from the Initial Tests

Whilst these RV series were unrealistic as they contained no stellar RVs, the results do indicate the vast differences of ability for the different schedules. Even in a 'perfect' world with only instrument measurement error and a simplified Keplerian model with no eccentricity, the reference schedules largely failed to detect the Earth-twin whilst the Terra Hunting and the Space schedules performed much better. The lack of false positives for the latter two schedules across the first 3 solar systems (bar one false for the 5-year Terra Hunting schedule on system2), indicates that the data coverage is adequate for low amplitude signal detections. It is interesting to see the difference in performance increase between the Reference and Terra Hunting schedules as the latter contains 3× the data but mostly improved on parameter estimation by many times more than $\sqrt{3}$. For the Terra Hunting, the occurrence of false positives is much less than the reference schedules, and the posteriors are much better constrained. For the space schedules, the complete coverage means the errors of the parameters likely follow a $1/\sqrt{N_{\text{data}}}$ relationship, i.e. if the data were increased by another factor of 2 (every 12 hours instead 24) the errors would likely only reduce by $\sqrt{2}$.

3.8 Second Results - Stellar RVs with a Constrained Model

In this section, I add the SOAP RVs and repeat the analysis. However, before assessing the rate of false positives, I firstly constrain PolyChord to only search the true model. In this scenario I explicitly tell PolyChord how many planets there are and allow it to search for the

optimum parameter values given the observation schedule. This initial test will allow the full unconstrained search in the next section to be put into perspective. As discussed in Section 3.6.4, the likelihood function and the model are *not* updated to include the variability from the stellar rotation effects. Here, I am investigating whether the sheer quantity of data can overcome this extra variability. The results are summarised in Table A.5.

3.8.1 System 1 Results - SOAP RVs, Constrained Model

For System 1 with the SOAP RVs included, see Figure 3.18b. Here, the single Earth-twin is now accompanied with the presence of attenuated RVs from spot groups rotating around the stellar surface, and a $1\sigma = 0.3\,\mathrm{m\,s^{-1}}$ measurement error on each data point. The true parameters of the Earth-twin are $K = 0.11\,\mathrm{m\,s^{-1}}$, $P = 293\,\mathrm{d}$, and $\phi = 3.46\,\mathrm{rad}$.

The 5-year reference schedule failed to confirm the parameters of the Earth-twin and instead found a higher amplitude $0.16\,\mathrm{m\,s^{-1}}$ signal at a shorter period of 210 d placed at the incorrect phase. Comparing the results with Section 3.7.1 where the schedule was able to find the planet, albeit with a large error, the presence of the stellar signals has clearly affected PolyChord's ability to find the planet with the incorrect likelihood function. The 10-year reference schedule performed better. It managed to find the planet but with large errors on the period and phase ($P = 302.69 \pm 65.13\,\mathrm{d}$, and $\phi = 3.01 \pm 0.44\,\mathrm{rad}$).

The Terra Hunting schedules both performed well and found good estimates on the RV semi amplitude and the period, but under-fitted the phase by a significant amount. The 5-year variant found a slight underestimate of the period of $P = 283.94 \pm 3.01\,\mathrm{d}$ but a larger discrepancy in the phase of $\phi = 2.44 \pm 0.22\,\mathrm{rad}$. The 10-year schedule found a better estimate of the period and phase of $P = 290.76 \pm 1.08\,\mathrm{d}$ and $\phi = 2.77 \pm 0.17\,\mathrm{rad}$. It is clear that PolyChord is able to find strong evidence of the true underlying signal, but that either the presence of the stellar variability from SOAP or the inadequate likelihood function is pulling the signal in phase-space. Either way, the results are promising from a Terra Hunting perspective.

The space schedules both found the Earth-twin with slightly better precision than the Terra Hunting schedules. For the 10-year variant, the phase of the planet is much closer to the true value with $\phi = 3.07 \pm 0.13\,\mathrm{rad}$ but it is still an underestimate. This result confirms that the stellar variability is affecting the fitting of a low amplitude pure Keplerian signal as this data set is continuous over the entire 10-year window.

3.8.2 System 2 Results - SOAP RVs, Constrained Model

For System 2 with the SOAP RVs included, see Figure 3.19b. Here, the Venus, Earth, and Jupiter-analogues are accompanied with the presence of RVs from spot groups rotating around

the stellar surface, and a $1\sigma = 0.3\,\mathrm{m\,s^{-1}}$ measurement error on each data point. The true parameters of the planets are: $K_{\mathrm{Venus}} = 0.11\,\mathrm{m\,s^{-1}}$, $P_{\mathrm{Venus}} = 197\,\mathrm{d}$, $\phi_{\mathrm{Venus}} = 3.46\,\mathrm{rad}$, and $K_{\mathrm{Earth}} = 0.11\,\mathrm{m\,s^{-1}}$, $P_{\mathrm{Earth}} = 293\,\mathrm{d}$, $\phi_{\mathrm{Earth}} = 4.45\,\mathrm{rad}$, and $K_{\mathrm{Jupiter}} = 10.34\,\mathrm{m\,s^{-1}}$, $P_{\mathrm{Jupiter}} = 2953\,\mathrm{d}$, $\phi_{\mathrm{Jupiter}} = 1.82\,\mathrm{rad}$.

The reference schedules both found the Jupiter-analogue with good estimates on the parameters. This is expected due to the sheer scale of the RV amplitude, see Figure 3.12 for an example of this. However, the schedules both found two low period false positives ($P = 49.20 \pm 28.09\,\mathrm{d}$ and $P = 27.50 \pm 16.94\,\mathrm{d}$) which are likely a result of the rotation period of the star of 25 d as this period had not previously appeared in any false positive until the SOAP data was included. The 10-year reference schedule may have a candidate for the Venus-analogue at $P = 198.67 \pm 175.24\,\mathrm{d}$ but the large error and incorrect phase could mean this is just a false positive.

Both of the Terra Hunting schedules performed better than the reference schedules. The 5-year variant was caught out by the presence of the SOAP RVs by finding a false-positive at $P = 20.88 \pm 2.05\,\mathrm{d}$ but found the Venus and Jupiter-analogues. The 10-year variant also found a false positive at $P = 80.38 \pm 33.64\,\mathrm{d}$ but this time found the Earth-twin and Jupiter-analogue with good estimates on all parameters.

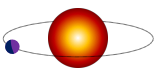
The Space schedules performed very well. They both found all three planets with good estimates on the parameters and small errors. The 5-year variant placed the Venus-analogue at a slightly larger phase of $\phi = 4.09 \pm 0.19\,\mathrm{rad}$ which could have been influenced by the SOAP RVs. The 10-year variant found all planet parameters within 1σ.

3.8.3 System 3 Results - SOAP RVs, Constrained Model

For System 3 with the SOAP RVs included, see Figure 3.20b. Here, the three terrestrial planets are accompanied with the presence of RVs from spot groups rotating around the stellar surface, and a $1\sigma = 0.3\,\mathrm{m\,s^{-1}}$ measurement error on each data point. The true parameters of the planets are: $K_1 = 0.16\,\mathrm{m\,s^{-1}}$, $P_1 = 101\,\mathrm{d}$, $\phi_1 = 3.46\,\mathrm{rad}$, and $K_2 = 0.13\,\mathrm{m\,s^{-1}}$, $P_2 = 197\,\mathrm{d}$, $\phi_2 = 4.45\,\mathrm{rad}$, and $K_3 = 0.11\,\mathrm{m\,s^{-1}}$, $P_3 = 293\,\mathrm{d}$, $\phi_3 = 1.82\,\mathrm{rad}$.

The reference schedules largely failed to find any of the true planetary signals and instead fitted 3 false positives. The 5-year schedule found evidence of a signal at $P = 203.33 \pm 25.85\,\mathrm{d}$ which could be the 2$^{\mathrm{nd}}$ planet, but placed it almost exactly π out of phase. The 10-year schedule found a false positive at $P = 342.78 \pm 133.52\,\mathrm{d}$ which could be an alias of the 365-day repeating schedule cadence. The other planets favoured here have large errors and periods that could be a result of the stellar rotation rate, e.g. $P = 51.32 \pm 24.55\,\mathrm{d}$.

The Terra Hunting schedules performed better and found the three terrestrial planets signals at the correct amplitudes, periods, and phases within 2σ. In particular the planets were well



constrained even in phase which is an improvement on the Earth-twin parameters found in the different systems. This could be by chance in that this particular phase sits well with the observation schedule cadence and the full signal has good phase coverage. To fully test this hypothesis, a range of identical schedules and planets would need to be assessed where *only* the phase is changed, and the posteriors examined for accuracy. It is likely that because the Terra Hunting cadence is 365-days and the planet period is 297-days, over a 10-year window there could be planetary phases that are preferential.

The space schedule also performed well. Interestingly, the 10-year space variant found the two outer Earth-twins but fitted the inner-most planet at a period of $P = 23.05 \pm 0.80$ d which is very likely to be a confusion with the stellar rotation rate. Otherwise the space schedules found good estimates with low errors on all of the remaining planets. Due to the continuous sampling of the Space schedule, it could be that this stellar rotation based false positive is favoured as this schedule is effectively sampling the stellar RVs extremely well, and the faint signal of the underlying planet is being ignored by the model. However in reality should such a data-set exist an observer would have an arsenal of stellar mitigation techniques to easily spot the stellar rotation rate and rule this signal out as a false positive.

3.8.4 System 4 Results - SOAP RVs, Constrained Model

There are no results to discuss for system 4 because I can not ask PolyChord for the parameters of a $N_p = 0$ model, I can only compare the evidence of this null model in the context of others. I discuss this in the next section.

3.8.5 Conclusions from the Constrained Search

With the presence of the SOAP RVs but constraining PolyChord to only search the true model, it is clear that the reference schedules still generally struggled to find the correct parameters of the planets and instead found a number of false positives for the multi-planet systems. A combination of a lack of data density and the under-sampling of the RVs are the likely cause of this as a couple of the false positives are on similar periods to the both of these.

The Terra Hunting schedules performed better and mostly identified the Earth-twin planets even when accompanied with others. This is early indication for the feasibility of the Terra Hunting Experiment as it is clear that the schedules have adequate data density to sample the 293-day signal, but more work would be needed to properly mitigate the stellar signals instead of the approximation I have made in this work.

The space schedules, predictably, performed the best. The continuous nature of the schedules gave PolyChord the best chance of fitting the models to the data, although it was still

struggling with System 3 as it found some false positives at stellar oscillation periods.

3.9 Third Results - Stellar RVs with an Unconstrained Model

In this section, the data fed into the PolyChord pipeline is the same as that in Section 3.8, but I now allow PolyChord to search all available models from $0 < N_p \leqslant 4$. With the benchmark set from the previous section, we are now able to assess whether PolyChord would actually prefer a different model from the forced true model, and if so what does it present as false positives or which planets does it omit as solutions.

3.9.1 System 1 Results - SOAP RVs, Unconstrained Model

For System 1 with the attenuated SOAP RVs included, see Figure 3.18b. The true parameters of the Earth-twin are $K = 0.11\,\mathrm{m\,s^{-1}}$, $P = 293\,\mathrm{d}$, and $\phi = 3.46\,\mathrm{rad}$.

Both reference schedule results strongly favour the $N_p = 4$ model with the $N_p = 3$ model being the nearest in significance in both cases and the true model being massively rejected with $|\Delta \ln R| = 19.39 \pm 0.29$ to the true model for the 10-year schedule. The parameters of the previously forced $N_p = 1$ model are no longer present and have been replaced with a set of different false positives. Both schedules find planets at integer multiples of the stellar rotation rate, e.g. $P = 25.63 \pm 16.65\,\mathrm{d}$ for the 5-year variant, and false positives at other periods also. The 10-year schedule finds a candidate with large bounds which do contain the true Earth-twin, e.g. $P = 425.90 \pm 181.64\,\mathrm{d}$, but the errors are so large that this candidate would likely be discarded.

The Terra Hunting schedules perform a little better although both results favour a $N_p = 4$ and $N_p = 3$ model for the 5 and 10-year variants respectfully. Both also massively reject the true model with the 5-year schedule find the evidence to be $|\Delta \ln R| = 25.70 \pm 0.28$ lower than the favoured model. Interestingly, the 5-year schedule finds a planet candidate that *could* be the Earth-twin whilst the 10-year schedule does not. The candidate mentioned has parameters $K = 0.13 \pm 0.02\,\mathrm{m\,s^{-1}}$ $P = 285.30 \pm 6.94\,\mathrm{d}$ which is promising, but again the phase is under-fitted at $\phi = 2.61 \pm 0.27\,\mathrm{rad}$. The 5-year schedule also finds 3 false positives that can likely be explained by the observation schedule cadence and the stellar rotation rate. The 10-year schedule may have found the Earth-twin with precise RV and phase estimates, but the errors on the period are large at $P = 453.28 \pm 343.07\,\mathrm{d}$. As mentioned before, shifting this planets phase relative to the observation windows would shed light on this hypothesis but this investigation did not extend this far.

Both of the space schedules favoured a $N_p = 4$ model but found the Earth-twins with well constrained parameters, but accompanied with 3 false positives. Particularly the 10-year

schedule found the period well at $P = 291.90 \pm 0.84$ d and had good estimates on the RV semi amplitude and phase also. This result indicates the power of data density and sampling even in the context of an incorrect likelihood function and model. The false positives for both schedules fall on short periods of less than 100 days and are at the bottom end of the RV semi amplitude prior indicating PolyChord is just fitting to the quasi-periodic signals in the stellar RVs.

The full results are displayed in Table A.6.

3.9.2 System 2 Results - SOAP RVs, Unconstrained Model

For System 2 with the SOAP RVs included, see Figure 3.19b. Here, the Venus, Earth, and Jupiter-analogues have true parameters of: $K_{\text{Venus}} = 0.11\,\text{m}\,\text{s}^{-1}$, $P_{\text{Venus}} = 197\,\text{d}$, $\phi_{\text{Venus}} = 3.46\,\text{rad}$, and $K_{\text{Earth}} = 0.11\,\text{m}\,\text{s}^{-1}$, $P_{\text{Earth}} = 293\,\text{d}$, $\phi_{\text{Earth}} = 4.45\,\text{rad}$, and $K_{\text{Jupiter}} = 10.34\,\text{m}\,\text{s}^{-1}$, $P_{\text{Jupiter}} = 2953\,\text{d}$, $\phi_{\text{Jupiter}} = 1.82\,\text{rad}$.

The 5 and 10-year reference schedules favoured the $N_p = 4$ and $N_p = 3$ models respectively, but the evidences were minimal for the $N_p = 3$ and $N_p = 4$ models respectively. The 5-year variant found the large Jupiter-analogue signal with moderate precision on all parameters, but failed to find any of the others. It may have found the Venus-analogue with a planet with $P = 168.51 \pm 117.21\,\text{d}$ at the correct phase of $\phi = 4.62 \pm 1.94\,\text{rad}$, but the large uncertainties place low confidence on the detection. The 10-year variant found a similar period signal but at an incorrect phase. Both found short period candidates that are likely inferred from the stellar activity.

The Terra Hunting schedules both favoured the $N_p = 4$ model with very strong significance, rejecting all other models strongly. Both found the Jupiter-analogue with good precision, and both have candidates for the Earth-twin. The 10-year schedule found the Earth-twin with good constraints on all parameters (particularly $P = 295.56 \pm 4.42\,\text{d}$) and a candidate for the Venus-analogue but it is placed at an incorrect phase. The 5-year variant returns results with increased uncertainties on the Earth-twin ($P = 321.02 \pm 42.21\,\text{d}$), and also found two false positives at stellar rotation periods.

The results of the Space schedules favoured the $N_p = 4$ models with even stronger significance than the Terra Hunting schedules. Both found all three planets with good estimates, the 10-year variant particularly excelled with the Earth-twin with $P = 293.23 \pm 1.53\,\text{d}$. They both also found an extra false positive at low periods similar to the stellar rotation rate. This is further evidence that there is an amount of data density that can overcome the stellar activity given a certain amount of stellar RV reduction. PolyChord was able to fit the Keplerian model to the planets and also to the periodicities found in the stellar activity, despite the white noise and SOAP RVs existing at much higher amplitudes than the signals of interest.

The full results are displayed in Table A.7.

3.9.3 System 3 Results - SOAP RVs, Unconstrained Model

For System 3 with the SOAP RVs included, see Figure 3.20b. Here, the three terrestrial planets have true parameters of: $K_1 = 0.16\,\mathrm{m\,s^{-1}}$, $P_1 = 101\,\mathrm{d}$, $\phi_1 = 3.46\,\mathrm{rad}$, and $K_2 = 0.13\,\mathrm{m\,s^{-1}}$, $P_2 = 197\,\mathrm{d}$, $\phi_2 = 4.45\,\mathrm{rad}$, and $K_3 = 0.11\,\mathrm{m\,s^{-1}}$, $P_3 = 293\,\mathrm{d}$, $\phi_3 = 1.82\,\mathrm{rad}$.

Both of the reference schedules favoured the $N_p = 4$ model with moderate to very strong evidence. Interestingly, the 10-year variant favoured the $N_p = 0$ model more than the $N_p = 1$ model which indicates that PolyChord finds that this under-sampled better describes pure Gaussian noise over a simple single Keplerian model. The 5-year variant found evidence for the Earth-twin at $P = 300.15 \pm 101.78\,\mathrm{d}$ but similarly large errors on the phase and RV semi-amplitude place this detection with low confidence. The rest of the planets found by the 5-year schedule are false positives with one being at the stellar rotation period. The 10-year variant also found a candidate for the Earth twin with large error on the period at $P = 338.09 \pm 112.05\,\mathrm{d}$ accompanied with 3 false positives.

The Terra Hunting schedules both favoured the $N_p = 4$ model with very strong significance. Both performed well in planet detection and found good estimates on all the parameters of the three planets. Each were accompanied with a false positive at the stellar rotation rate. In particular the 10-year schedule found $P = 291.99 \pm 0.93\,\mathrm{d}$ at a phase of $\phi = 1.68 \pm 0.14\,\mathrm{rad}$.

The Space schedules also favoured the $N_p = 4$ models with very strong significance. Interestingly the 5-year variant was out performed by the 5-year variant of the Terra Hunting schedule for the first time as it found the outer two planets accompanied with two shorter period false positives. The 10-year schedule found the three Earth-twins with good estimates on all parameters, and a false positive at the stellar rotation rate.

The full results are displayed in Table A.8.

3.9.4 System 4 Results - SOAP RVs, Unconstrained Model

For System 4 with the SOAP RVs included, see Figure 3.21b. Here, there are no planets and so no true parameters to find and hence any planet found is a false-positive by definition.

The reference schedules were the only ones to successfully identify the correct model of $N_p = 0$. The 5-year variant only moderately favours this over all other models whilst the 10-year schedule massively favours the true model over all others. This is a rare moment of success for these schedules or it could be that the lack of sampling and large gaps force PolyChord to favour nothing over a Keplerian signal. This model selection does *not* mean that PolyChord thinks the signals are stellar noise and therefore rejects the Keplerian model, it only means that PolyChord has not favoured the Keplerian model.

The Terra Hunting schedules favour $N_p = 3$ and $N_p = 4$ for the 5 and 10-year schedules

respectively with moderate to strong significance. The false positives in each come from the stellar oscillation signals and the 180-day repeating window of observation cadence. Other, longer period false positive planets could be a result of a combination of these.

The Space schedules favour the $N_p = 4$ and $N_p = 3$ planet models for the 5 and 10-year schedules respectively with strong confidence. Like the Terra Hunting schedules, the false positives sit on periods similar to the stellar rotation rate. The longer periods of the false positives may be harmonics of the spot group lifetime and the rotation rate but more investigation is required to confirm this. All of the RV semi amplitudes for these false positives are at the lower end of the RV prior, indicating that PolyChord has essentially found some quasi-periodicity in the data but is fitting it to noise.

The full results are displayed in Table A.9.

3.9.5 Conclusions from the Unconstrained Search

The unconstrained search provided more insight into the success and failures of the various schedules. Across all types, false positives relating to the stellar rotation rate were found, which is unsurprising giving the lack of stellar modelling. The Terra Hunting and Space schedules performed well and found the Earth-twin candidates with moderate success, with and without companions. This is an early indication that the level of data density predicted by the Terra Hunting Experiment is sufficient for Earth-twin detection.

The fact that the Terra Hunting and space schedules did not identify the null-case of zero planets is also unsurprising given the amount of data. In Figure 3.13 the stellar rotation rate can be clearly seen for these schedules but not for the reference ones, so a Keplerian model fitted here would have a larger evidence than no model at all.

The planet search was limited to 4 in each case to save on computational expense. It is likely that if this limit was lifted, more false positives would be favoured in each case.

3.10 Examining the Results in More Detail

As PolyChord returns the full posteriors for each parameter of each model tested, the like-wise comparison can be made between a given parameter of two different schedules. The values presented in the previous section are the weighted average of the posterior with the error on the weighted average calculated from the distribution. But these values do not tell the whole story. A user can check if the ideal Gaussian nature of a posterior is actually multi-modal or a more complicated shape which would indicate multiple plausible solutions for that model.

Traditionally, corner plots are used to show correlation between a pairs of a multi-dimensional posterior space, such as those in Figure 3.17. However, for large dimensionality

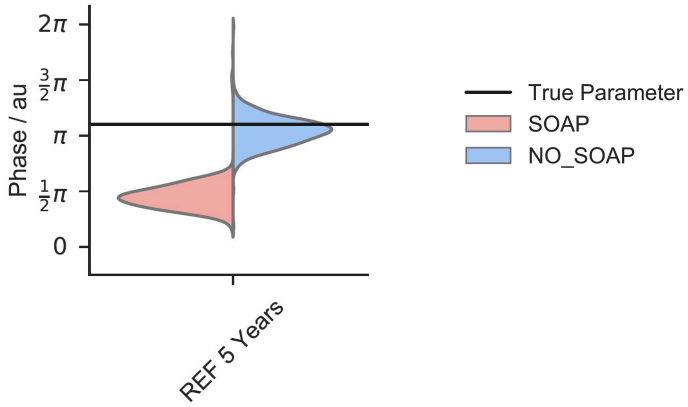


Figure 3.22: An example of a violin plot for a single parameter. The two posterior distributions are of the orbital phase of the planet for the 5-year reference schedule (REF 5 Years). The left (red) distribution is the result when the SOAP RVs are included, the right (blue) distribution is when the SOAP RVs are omitted. The horizontal black line is the true value of the parameter.

they are cumbersome to read and difficult to interpret. Also, in this investigation there are two sets of posteriors to examine, the case where the SOAP data is present and where it was omitted. I chose to present my results with in the form of a Violin plot from the SEABORN package[i]. An example of how to interpret a violin plot is displayed in Figure 3.22. Here the orbital phase of a planet is shown and a pair of distributions with different widths and means are plotted. A violin plot is an excellent way of comparing two parameters that are related but differ as per some type or label, in this case the stellar oscillations from SOAP.

In Figure 3.23, I present the results from Sections 3.7.1, and 3.8.1 where I forced POLYCHORD to the correct model for System 1. Here, the six observation schedules each are associated with a posterior per parameter with and without the stellar signals from SOAP. This enables the quick and efficient comparison between these two data sets across all parameters and schedules. There are a number of distinguishing features of this visualisation, firstly the inclusion of the SOAP data has clearly affected each schedule by a measurable amount. The reference schedules appear to be more affected than the Terra Hunting and Space schedules. Also, as the number of data points increases from left to right, so do the posteriors accuracy on the true value and their width decrease. For the 10-year Terra Hunting schedule and the 10-year Space schedule, the inclusion of the stellar data has had little effect on the RV semi amplitude, but has affected the period and the phase.

In Figure 3.24, I present the results from Sections 3.7.2, and 3.8.2 where I forced POLYCHORD to the correct model for System 2. In this figure, there are 3× the posteriors as compared to Figure 3.23 as we are now comparing the results from 3 planets. For planet 1, the Venus-analogue, the

[i]The package information can be found here: https://seaborn.pydata.org.

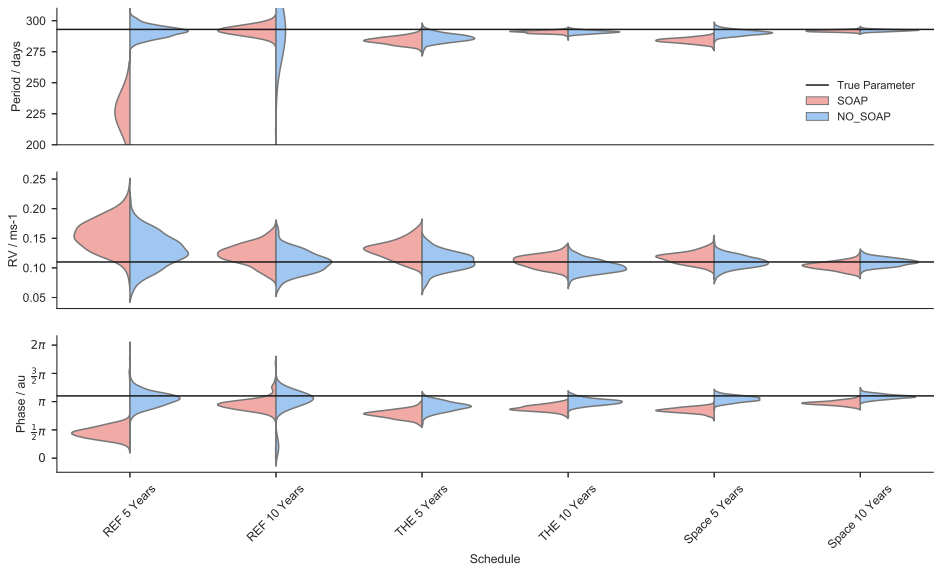


Figure 3.23: The posterior distributions for the favoured solutions of the 1-planet model for the various schedules observing system 1 with and without the SOAP RVs included. See Figure 3.22 for explanation.

posteriors reported in the previous sections are visualised. The reference schedule posteriors are broad and multi-modal whilst the Terra Hunting and Space schedules are narrower and Gaussian and are closer to the true value. The inclusion of the stellar RVs from SOAP has clearly affected the reference schedule's ability to fit the phase, but does not affect the others to the same degree. For the Earth-twin, the 10-year Terra Hunting and both space schedules perform comparably with neat Gaussian posteriors. For the Jupiter-analogue, again the last three schedules perform the best, whilst the reference schedule only succeeds without the SOAP RVs, and even the posteriors are very wide.

In Figure 3.25, I present the results from Sections 3.7.3, and 3.8.3 where I forced PolyChord to the correct model for System 3. The inner-most Earth-twin posteriors are reasonably Gaussian for most of the models and the SOAP data has not affected the results as much here. It could be the relative periods and amplitudes of the the SOAP RVs and this particular planet are easily discernible. The Venus-analogue posteriors are, again, more Gaussian in nature than those found for System 2. The SOAP RVs have had more of an effect here and have shifted most of the 'No SOAP' results off from the true value. Lastly, the Earth-twin was well placed by the Terra Hunting and Space schedules in phase-space, but the presence of the SOAP RVs clearly lead to an overestimation of the planets RV Semi Amplitude. This shows the vulnerability of this technique as the mass-estimate of the planet would be off by almost 50% for these detections.

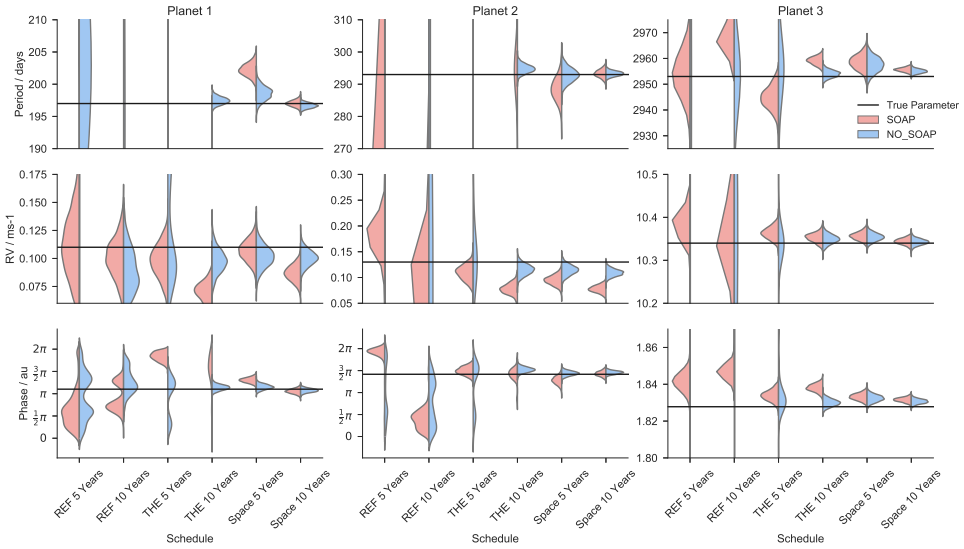


Figure 3.24: The posterior distributions for the favoured solutions of the 3-planet model for the various schedules observing system 2 with and without the SOAP RVs included. See Figure 3.22 for explanation.

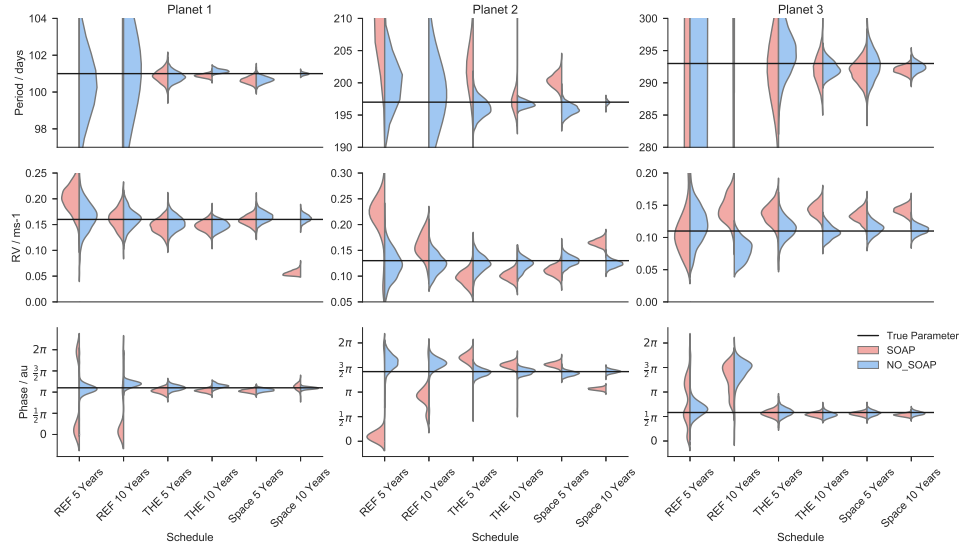


Figure 3.25: The posterior distributions for the favoured solutions of the 3-planet model for the various schedules observing system 3 with and without the SOAP RVs included. See Figure 3.22 for explanation.

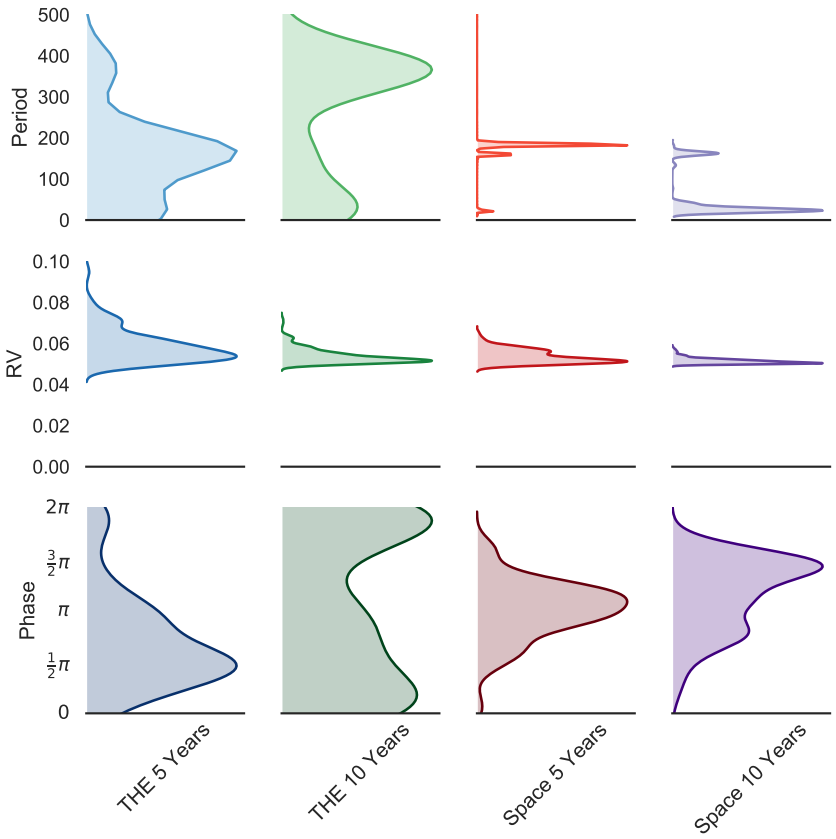


Figure 3.26: An example of some of the posteriors favoured for system 4 for the schedules that did not confirm the true $N_p = 0$ model. See Figure 3.22 for explanation.

Lastly, for Figure 3.26, a set of posteriors favoured by the Terra Hunting and Space schedules are shown. Both of the reference schedules correctly identified that this is the $N_p = 0$ model and so there are no posteriors to show. Here, it is clear that the RV Semi Amplitudes are pushed to the lower-end of the prior, whilst the periods are of stellar rotation rate and 180-day based scheduling. The phase is almost uniform which also suggests that PolyChord could fit some periodicity, but the quasi-periodic nature meant there is no dominant amplitude or phase. A likelihood function or model incorporating this kind of variability may detect this, and in that case these schedules may then correctly identify the $N_p = 0$ model.

For a visual inspection of the PolyChord solutions, I can take the peak of the posterior of the period parameters and phase-fold the RV curve on that period. This is a traditional visualisation used in many exoplanet discovery papers as it allows a long series of data to be focused around a particular event such as a transit.

In Figure 3.27 I have taken the data from Section 3.7.1 which is the System 1 RVs with only Gaussian noise as sampled from the 10-year Terra Hunting schedule. For this system, I

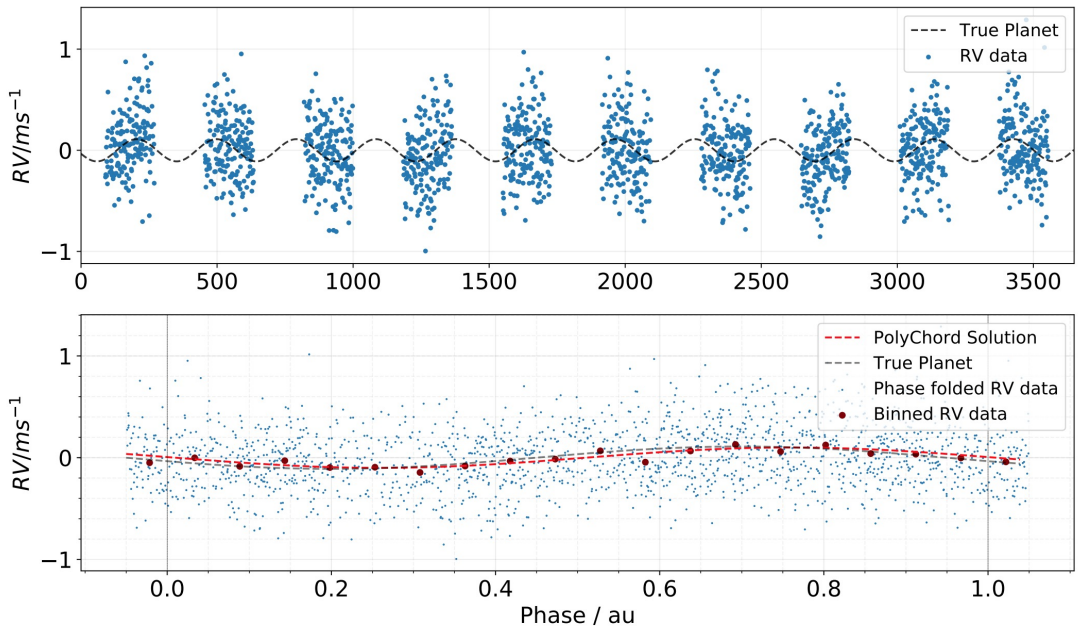


Figure 3.27: The original (top) and phase–folded (bottom) time series of 10-year Terra Hunting schedule observing System 1 with only Gaussian noise. The solution favoured by PolyChord is drawn on the phase–folded plot as the red dotted line whilst the true planet is the black dotted line.

retrieved the favoured period of the favoured 1-planet model and phase folded the data. I have also superimposed the true planet and drawn in the PolyChord solution to show how visually close they are.

In Figure 3.28 I have plotted the data from Section 3.8.1 where the data now contains the SOAP RVs. As before, I phase folded that data on the favoured period value, and generated an RV curve from the rest of the parameters returned by PolyChord. Here, the slight underestimation of the phase due to the presence of the stellar activity is more visible.

I also used a traditional frequency analysis with a Lomb-Scargle periodogram to examine periodicities in the data (Lomb 1976; Scargle 1982). This is an approach traditionally used by the Exoplanet community to find sinusoidal signals in data but can be less trusted when the signals are well below the noise level or are poorly sampled (VanderPlas 2018), as in the case for the Terra Hunting Experiment. Figure 3.29 shows the Lomb-Scargle periodogram for five different time series to assess any inherent periodicities in the data. The first two are the 'binary masks' of the ground-based schedules, i.e. a '1' when there is an observation and a '0' when there isn't. Because this signal is essentially a set of top-hat functions of equal amplitudes, the Lomb-Scargle will pick up any repeating patterns with no bias to their amplitude. Then, I show the time-series for the System 4 data sets observed by the three different 10-year long

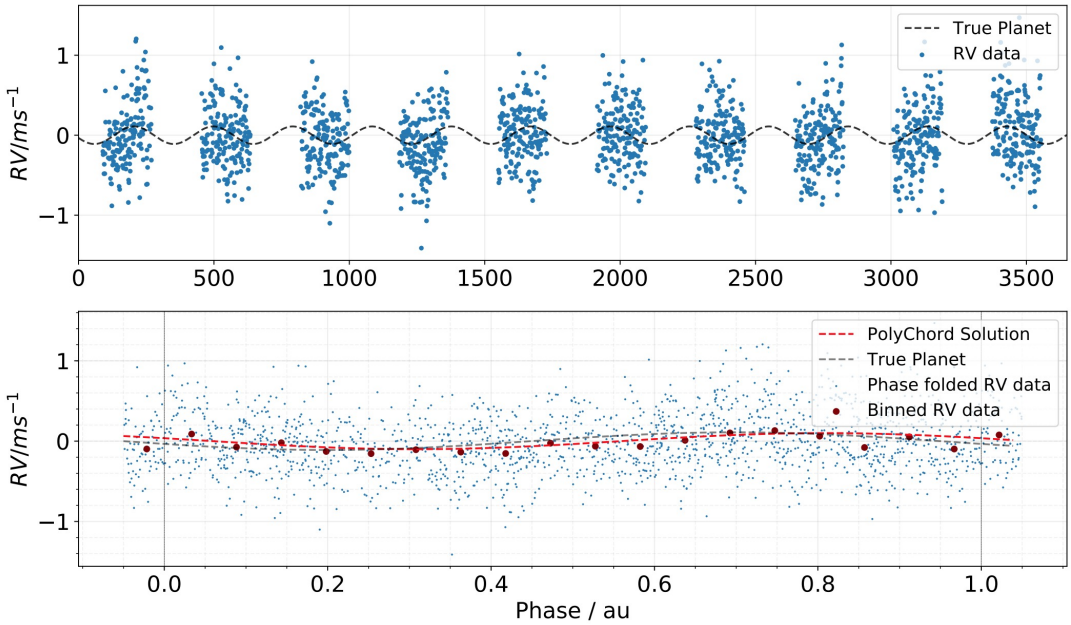


Figure 3.28: As per Fig. 3.27 but with 10-year Terra Hunting schedule observing System 1 with both Gaussian and SOAP noise.

schedules. For each, I have computed the Lomb-Scargle periodogram and plotted different false-alarm probability (FAP) levels. The FAP is the probability that a given peak is or is not due to random error, given the assumption that the noise on the data is purely Gaussian. E.g. for any peak above the 1% FAP line, there is a 1% chance this peak is erroneous, hence lower amplitude peaks are generally less trusted.

In Figure 3.29 there are a number of significant peaks especially in the uninterrupted SOAP RVs (equivalent to the 10-year Space schedule observing system 4). Not surprisingly, there is a strong peak at the stellar rotation rate of 25 days. The ground based schedules also show significant peaks at integer multiples of their own respective window function, e.g 10-days, 30-days, 180-days, 365-days. Not that this work is the first to show it, but this type of analysis demonstrates the need for modelling over pure frequency analysis, and it demonstrates the need for accurate models that describe the data as best as possible.

3.11 Conclusions and Future Work

In this chapter I have presented a description of a new RV analysis pipeline that incorporates nested sampling with a Bayesian framework for model comparison. The literature review surrounding the stellar activity highlights the necessity for accurate modelling of not only the Keplerian planetary signal, but of the stars themselves.

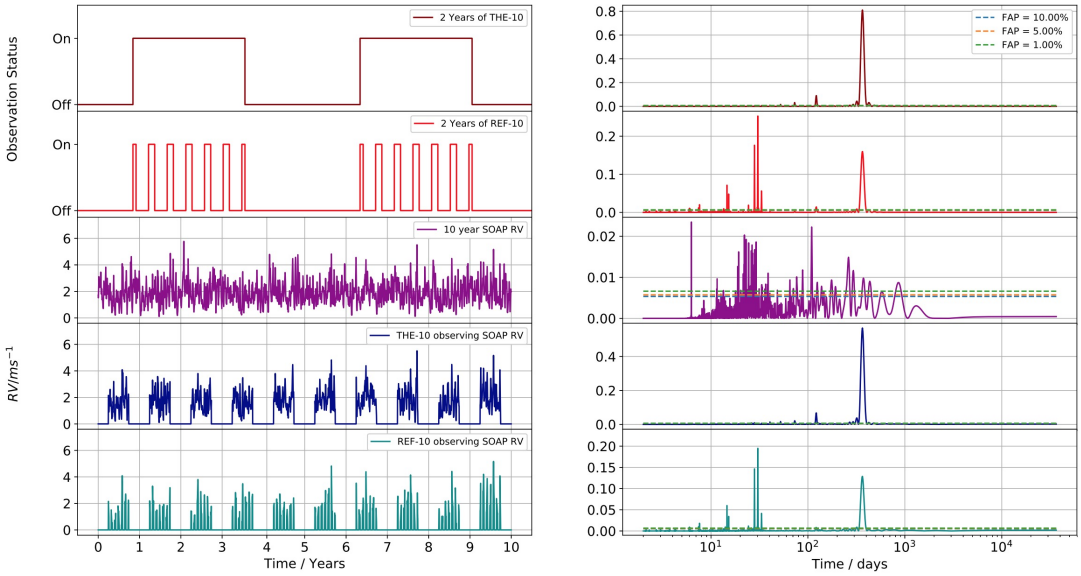


Figure 3.29: Frequency spectra of the observation schedule's window functions and of the stellar signals from SOAP observed by the schedules. From top to bottom, each pair of plots (left and right) are a time series and the Lomb-Scargle periodogram. In order they are the binary mask of the Terra Hunting schedule, the binary mask of the Reference schedule, the uninterrupted SOAP RVs as seen by the space schedule (System 4 data), System 4 as seen by the Terra Hunting Experiment, and System 4 as seen by the Reference schedule. False alarm probabilities have been plotted on the periodograms at 10%, 5% and 1% for each.

I have used this pipeline to simulate a set of RVs stemming from different solar systems observed by different schedules, and have fitted various planetary models to each. The pipeline produces model comparison and posterior estimation for all parameters of all the models such that a user can examine even the parameters of an unfavoured model to further their investigation of the data.

With this framework in place, I conducted an initial feasibility study for the Terra Hunting Experiment to answer the seemingly benign question of 'Can the Terra Hunting Experiment find an Earth-twin in the habitable zone of a Sun-like star?'. I have demonstrated, via the comparison of various schedules and solar systems, that the Terra Hunting schedule outperforms a typical ground-based RV survey on a near identical instrument, and performs comparably to a continuous schedule with high level of stellar modelling to reduce the stellar activity RVs.

There are a number of avenues that this research can now take, a few of which are currently (at the time of writing) being investigated within the Cavendish Exoplanet Group. A first and obvious step is to include stellar variability into the likelihood function as seen in Feroz et al. (2011). Even a simple linear parameter can account for quasi-periodic variability which could drastically change model selection as this work often favoured a high number of planets with

false positives at stellar rotation rates.

To take this further, the user could also examine the time-series data from the BIS, FWHM, and the FF'. Any periodicities found here could indicate the presence of stellar activity, which could then be fed into the priors of a purely Keplerian model by reducing the prior value at these identified periods. This would limit the effect of the stellar rotations as PolyChord would tend to reject signals at these frequencies, but it could also have the side effect of missing true positives near these periods.

Also, the simulated stellar variability from SOAP2.0 misses a key activity cycle: magnetic activity. SOAP generates activity at a consistent level which is therefore inadequate for a 10-year survey simulation as we know from the Sun the activity level will have passed almost one complete cycle in this time. A quick-fix for this could to be to modulate the SOAP data with a variable term of a determined size and period, and to re-run the analysis. For example, a quick but poor approximation would be to multiply the SOAP values by a sinusoidal term. Another option could be to look at the sunspot numbers over a 10-year period and force SOAP to generate similar numbers of spots or have the probability of generation being proportional to this curve. Alternatively, the SOAP source-code itself could be adjusted to factor in this long-term oscillation with a more physical origin.

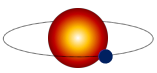
Also, the HARPS-N Solar Telescope has now over 3-years of RV data from the Sun as a Star by directly imaging the Sun (Haywood et al. 2016). When reference-frame corrected, this RV series acts as an ideal substitution for the SOAP data but it currently covers just over a 3-year period. It may be possible to extrapolate this data, using sophisticated methods such as a trained Gaussian Process or a machine learning algorithm, into a 10-year dataset to be inserted into this type of simulation.

For the ground and space-based cadence, I was somewhat forced to use a 24-hr time-stamp for each observation due to the constraints from SOAP. In reality, an observation schedule is more dynamic due to local constraints and other targets scheduled for that night. At the time of the simulation, the Terra Hunting Experiment scheduler was still under construction but it is now at a stage where it can output a predicted set of time-stamps for a complete 10-year survey. This not only allows for the realistic sampling of targets from the ground, but for a full yield-estimation for a multi-target list. Each target could be accompanied with a unique data-set from SOAP, and given a planetary system generated from known distributions of planet occurrence rates for that type of star. Then, a complete analysis of the *entire* schedule performance could be conducted. Similar work, such as for the NGTS (Next Generation Transit Survey) (Günther et al. 2017), has been instrumental in assessing the yield rate and false positive rate for long-term exoplanet surveys.

However, for this level of analysis the PolyChord code would need to be drastically sped

up. A current Masters student in the Cavendish Exoplanet Group has taken my code and re-written it in C++ for speed and is observing 50-100 fold reduction in analysis time for similar sized datasets. This work is also conducting an investigation in including stellar activity in the likelihood by combining several models in parallel that look at the various time series from HARPSN data (RV, BIS, FWHM, Flux), and is back testing the new analysis on previously confirmed planets (Ahrer, Hall et al. in prep).

Whichever the direction taken, the potential results for this line of work are vast, and the foundation currently set provides hope for long-term ground-based surveys aiming for Earth-twin discoveries.



Chapter 4

The Detector Technology used for Exoplanet Discoveries

'Without stories, we wouldn't be human beings at all'

Phillip Pullman

High resolution optical detectors have been a key development for the history of astronomy. They allow the few precious photons that reach the Earth to be collected in a digital format for scientific evaluation and assessment and, crucially, can be used in a predictable and repeatable manner. These detectors are a key consideration when designing new instruments and missions as the choice influences key parameters such as photometric precision, spectral resolution, and exposure time. In this brief chapter I will outline the basic principles of optical detectors and in particular Charge-Coupled Devices (CCDs). I will then discuss some standard calibration experiments that are routine for a new optical system and finish with a justification of the need for per-pixel geometry characterisation for Earth-twin discoveries via the radial velocity technique. A large portion of my understanding of the history and functionality of CCDs comes from the highly regarded textbook of Janesick (2001), and I encourage the interested reader to start here when embarking on any project that involves them.

A CCD is a semiconductor device constructed from an array of small light-sensitive cells known as pixels. Should a photon fall onto one of these cells, there is a chance that it will be converted into an electron due to the doped nature of the underlying material which is usually

Silicon for optical wavelengths due to the band-gap energy of the material, but sometimes can be the alloy of Indium-Gallium-Arsenide for red to infrared applications.

The concept for the underlying technology behind CCDs was developed in the late 1960's in Bell Labs by Willard Boyle and George E. Smith[a] in their investigation of shifting charge from one register to another. The first crude imaging devices were only a handful of pixels in size, and used the energy of the photons to build up charge in the storage capacitors before they were clocked out by a simple register (Tompsett et al. 1970). Within a few years the technology had rapidly developed into a square array of many hundreds of pixels that could operate at high frequencies with integrated optics to focus the image. Since then, development has been mostly focused on increasing the resolution, well-depth, readout speed, and in reducing the noise. Modern CCDs for astronomy, often the highest quality manufactured, are many thousands of pixels in size on a side, and have well-depths of over 100,000 electrons. Boyle and Smith shared $1/2$ of the 2009 Nobel prize for their contribution to CCD technologies.

4.1 Basic Functionality of a CCD

The photosensitive region of the CCD, the pixel, will passively build up charge as more photons promote electrons into the conduction band of the material. The charge builds up until the potential well of the pixel is full. A positive voltage applied to the gate above the pixel will attract all of the accumulated electrons. Voltages are applied to the gates in such a sequence that allows the packets of electrons to be shifted across to the next pixel without interacting. Once the whole row is clocked out to the register, each packet is counted and that number is stored digitally as the counts of that pixel and the next row is then clocked out. Modern CCDs can achieve upwards of 5 MHz in frequencies and a large-format CCD can be read-out in seconds. Figure 4.1 shows this sequence for 3 neighbouring pixels. The insulating layer prevents charge leakage vertically whilst the voltage of the gates prevents charge leakage laterally.

The clocking sequence is easier said than done. Bespoke CCD controllers are required for large format astronomy CCDs (e.g 4096×4096 pixels) as the timings on the voltage signals are often sub 1 µs and are assumed to be 100% reliable for millions of readouts. For less precise applications, such as hand-held single lens cameras, CCDs can be clocked with smaller and cheaper circuitry however this often increase the noise level and count errors.

As CCDs are passive in generating photoelectrons, the pixels will continue to register photons even as they are being clocked out. For a large CCD this means that the very first pixel could have much fewer counts than the last pixel, as this last pixel will be sitting there waiting for the voltage signals to transfer the electrons. Typically, this is addressed with the use of a

[a] See patents US3792322A and US3796927A.

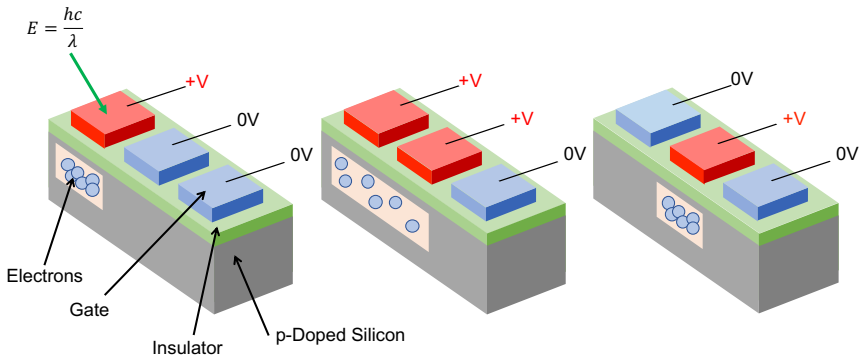


Figure 4.1: The clocking sequence of a 3×1 pixel CCD. The photon hits the surface and charge is generated proportional to the exposure. Careful timing of applying a voltage in sequence will move the charge along the row as a packet.

shutter, the CCD will be enclosed in darkness and only exposed to the source light for a fixed duration as determined by the shutter control.

4.2 Characterising a CCD for Astronomy

In this subsection I detail a few typical procedures that are used to characterise a CCD to be used for astronomical imaging, and list a few of the common sources of noise and error.

4.2.1 Dark Current

CCDs operated at room temperature are subject to the build up of counts registered from thermal excitation within the semiconductor layers themselves. These counts are known as 'dark current' as they appear as a count even in the absence of light. They arise from electrons who spontaneously promote themselves to the conduction band due to latent thermal energy in the material. This process is random in nature but scales with temperature. Thus, CCDs are generally operated at low temperatures to reduce this signal to just a few counts per exposure over the entire CCD. Typically thermoelectric cooling can reach −80 °C whilst cryogenic cooling with liquid nitrogen can reach −200 °C. To measure the dark current, exposures of a fixed time in a blacked-out enclosure are taken at various temperatures. The counts will be seen to gradually reduce until at an acceptable level as determined by the science goals of the instrument.

4.2.2 Read Noise and Bias

During the clocking process, errors in measuring, amplifying, and setting the voltage (bias) can cause an incorrect count to be measured. This process contributes another source of normally distributed noise to a typical read-out process but can be easily corrected for with an exposure of

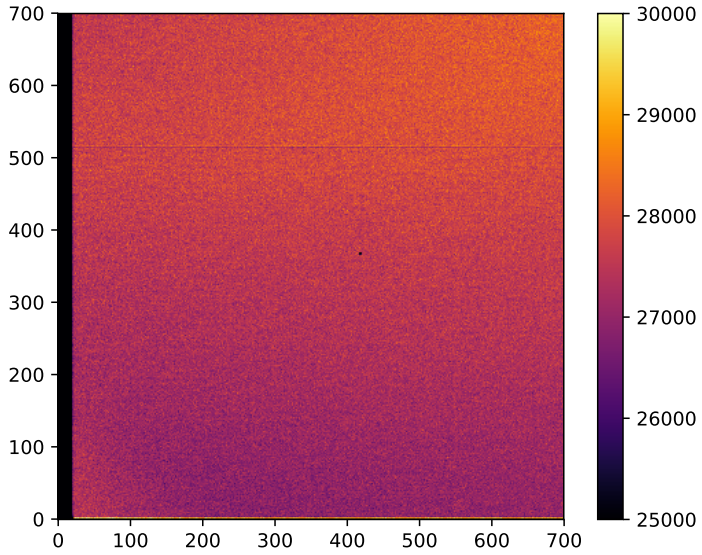


Figure 4.2: A 700 pixel square crop of a flat-field test from one of the SPECULOOS CCDs. The dark horizontal bar is at pixel 512 is the effect of a slightly misaligned mask during fabrication. This results in that row being slightly too large or too small and will collect more or less light relative to the rest of the pixels. The colour range is chosen to exaggerate this feature. Also visible on the left is the over-scan region that is used to measure the read-noise of the detector, Section 4.2.2, and a dead pixel at ~(420,380).

0 s in a completely dark environment. Modern CCDs usually have a virtual set of extra columns either side of true imaging area to measure this signal. Because these pixels do not actually exist, the counts measured are simply a random process of latent electronic noise that will be common to the imaging section also. The zero exposure time allows for the differentiation between the bias and dark current.

4.2.3 Flat-field Images

Despite the efforts in semiconductor manufacturing, a pixel may still register a slightly different number of counts than a neighbour even under identical circumstances. Generally, large-scale structure or gentle gradients can be seen which are usually a consequence of non-uniformities of the underlying material. The scale of this structure can be just a few counts in 100,000, but this can be a significant discrepancy for astronomy. To mitigate this, the CCD is illuminated with a known uniform field of light (hence the term flat field) and the photometric response is taken. Flat frames are then used in data pipelines to correct sources of light at different points on the detector e.g. a pair of stars in different corners may have the same brightness but could

register different on a uncorrected CCD.

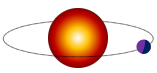
The flat field also reveals dead pixels, stitching boundaries, charge transfer issues and other minor defects. Flat field images can also be used to verify the linearity of a detector. Should a given intensity register 100 counts on the detector, an astronomer will hope that doubling the intensity or the exposure time will result in 200 counts. If so, they can reliably conduct research on a variety of objects in their field of view, and measure photometric signals with accuracy such as a transiting exoplanet. Most CCDs are linear for a majority of their well-depth capacity but can drift from linear at the lower and upper few % of saturation. Hence, knowing the linearity response allows correct exposure times to be used to ensure the data is in the linear portion of the detector.

The flat field can also reveal another seemingly photometric noise source that is a result of the fabrication of large format CCDs. A photolithography mask contains the image of a block of pixels, say 1024x512, which is then stepped and repeated across the silicon wafer to create a detector of a given size. However, slight misalignments even of just 100 nm, will cause the pixels of the joining rows to be larger or smaller than desired. This means for a perfect flat-field illumination, it will register more or fewer counts than expected because it will collect more or fewer photons due to its size difference. Hence, a flat field can reveal some *geometric* discrepancies of the detector that can severely impact exoplanet searches. Figure 4.2 shows a subsection of a flat-field image from a SPECULOOS CCD[b]. Apart from the pixel to pixel brightness variation, there is also a stitching boundary present in the horizontal line across the detector.

4.2.4 Quantum Efficiency

The quantum efficiency (QE) is the measure of the percentage of photons that actually get converted into electrons as a function of wavelength. Different coatings and materials laid on the CCD surface can selectively improve or reduce the QE as desired, along with variations of material thickness and doping. Typically, science-grade CCDs have a QE of over 75% in the desired spectral range. Figure 4.3 shows the QE from the CCD to be used for the HARPS3 CCD tests, this detector is quarter-sized as compared to the science-grade device that will actually be used in the instrument as it is 2048×2048 pixels compared to 4096×4096 pixels in size.

[b]SPECULOOS is a ground-based transit survey targeting M-dwarf stars. During the early phase of my PhD I worked with these detectors during their final commissioning phase to practice the various techniques of scientific CCD characterisation.



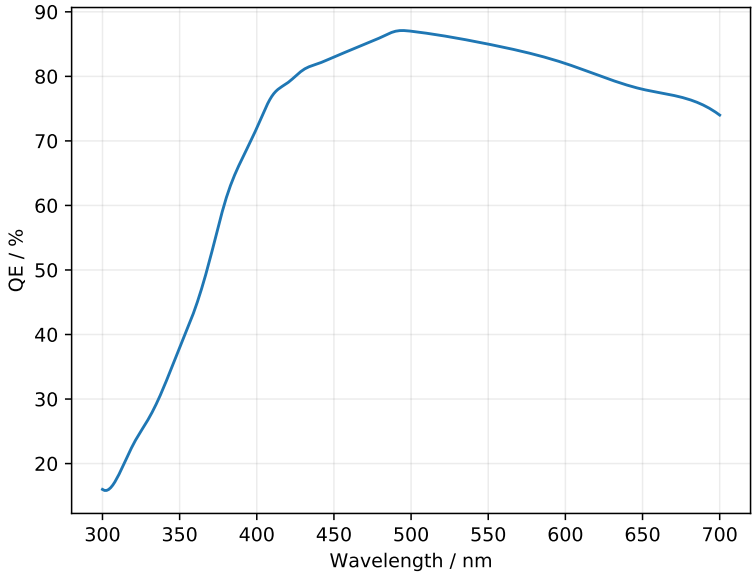


Figure 4.3: A Quantum Efficiency Curve. This QE curve shows the device is much more sensitive in the green and red end of the visible spectrum as opposed to the blue end. This particular device is the one used in the tests discussed in Chapter 5.

4.2.5 Shutter Map

Whilst the shutter solves the issue of non-uniform charge building up during the readout, it also introduces its own source of photometric variation that needs to be accounted for. For a completely flat field of light, the shutter opening and closing will mean that each pixel receives a different number of photons. A shutter is usually a multi-vaned mechanism that takes a non-zero time to fully open and close. During the opening of a shutter, the central pixels are exposed first whilst the edge pixels are still under the shadow of the shutter. Then when the shutter is closing, the edge pixels are cast under shadow first with the central pixels the last to be blacked out. This has the effect of imprinting the shape of the shutter on the image, see Figure 4.4. Modern shutters can open and close in just a few ms, but this is still long enough to have an effect on shorter-exposures of under a 1 s. A shutter map is a per-pixel function measured from the combination of flat-fields and shutter timings that relates the amount of flux a pixel receives to the position it is on the detector relative to the shutter. Each detector and shutter combination will have a unique shutter map. This map is then used to transform the image back into a true flat field. It's worth noting here that a shutter is not actually used in a spectrograph as the exposures are often as long as 20 minutes and the shutter open/close time negligible on this timescale, but one will be used in the experiment to measure the pixel positions of the CCD.

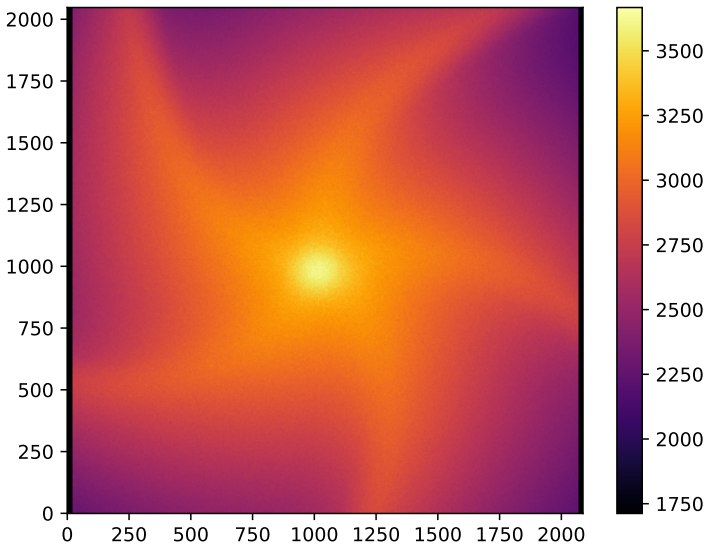


Figure 4.4: The Shutter Map produced from a 0.01 s exposure from one of the SPECULOOS CCDs. The short exposure time exaggerates the effect of the shutter as it is comparable to the time it takes to open the shutter.

4.3 Why CCD Pixel Mapping is Needed

Aside from the stitching boundaries found with a flat-field, all of the calibration techniques mentioned in the previous section are *photometric* calibration techniques. I.e. they are used to characterise the photometric response of the detector under different circumstances to better understand the inherent behaviour. However, the *geometric* variation of the pixels of the detector will influence how we interpret data. In the case of monitoring spectral motion for exoplanet discovery, an uncharacterised detector may lead to time-dependent false positives propagating through an analysis procedure and, ultimately, false planet detection.

In Dumusque et al. (2015b), the authors noted a 'spurious' 1–year signal in many data sets with an amplitude up to a few m s^{-1}. Suspecting these signals were local rather than from a number of exoplanets at exactly 1–year periods, they found that in these measurements, a certain spectral line was crossing a specific stitching boundary due to the motion of the Earth towards and away from the target. Because the data reduction pipeline did not factor in this row with larger pixels, the correction to the spectra from the Earth's motion was incorrect and therefore introduced a 1-year period signal. When this line was removed from the overall analysis of the CCF, the 1-year signal disappeared from the analysed RVs.

The stitching boundaries are fairly trivial to mitigate. Usually, the affected pixels are known

from the manufacturing data supplied with a detector, e.g. a 1024×512 pixel block, so it is possible to design the spectrograph optics axis of rotation to ensure the orders cross the fewest boundaries. Or, as in the case detailed above, it can be treated in the spectral analysis.

But, it is not just these pixels that have displacement errors. The production quality of the actual mask used during photolithography will have some inherent random and systematic position errors. Each pixel is likely to be *nominally* centred about its intended position, but it could deviate by some small fraction of a pixel. Also, if the step and repeat positioning of the mask is not precise, the entire block could be translated or rotated about its intended position which would effect all of of the pixels of that block and not just those at the edge.

The order of magnitude of the displacement is likely to be small for the sub-block displacements. The deviations within the photomask are likely to be precise to a few nanometers ($\sim 1 \times 10^{-3}$ of a pixel) and the block stitching errors are on the scale of many tens to a hundred nanometers, all much smaller than a typical pixel side length of 15 µm. As discuss in Section 1.2.1, the predicted motion of a spectral line of HARPS3 for an Earth-twin scale RV is around 1 milli-pixel, which is 15 nm in absolute terms on the detector. Hence, the objective of the experiment is to characterise the pixel positions of an optical detector to this level.

4.3.1 The RV Error Budget

To estimate the overall contribution to the error from a non-mapped detector, a comprehensive instrumental error budget must be computed. This is a typical procedure for an astrophysical instrument during the final stages of proposal, and are often used as benchmarks during construction, installation, and testing of the final product.

For an RV spectrograph, the RV error budget comprises all the different sources that can contribute to the error bar of a given data point, and is therefore used to assess which parts of the instrument or data analysis pipeline contribute the most. For HARPS3, the predicted error budget sums to $0.3\,\mathrm{m\,s^{-1}}$, this value stems from estimates of the current performance of HARPSN, and the predicted improvements of the newer design and data reduction pipelines (Thompson et al. 2016). An in-depth study into the full component's contribution is yet to be conducted.

However, there is an instrument that is currently undergoing the final stages of construction and testing whose error budget can be used as a model for HARPS3. The NEID[c] spectrograph has a similar mission to that of HARPS3 and the Terra Hunting Experiment, *and* considered the effects of a pixel-mapping procedure in the context of a global RV error budget. In Halverson et al. (2016), the authors factored in a variety of sources of error and categorised them into

[c]Pronounced 'noo-id' or 'new-id', the word means 'to see' in the native language of the Tohono O'odham, on whose land Kitt Peak National Observatory is located.

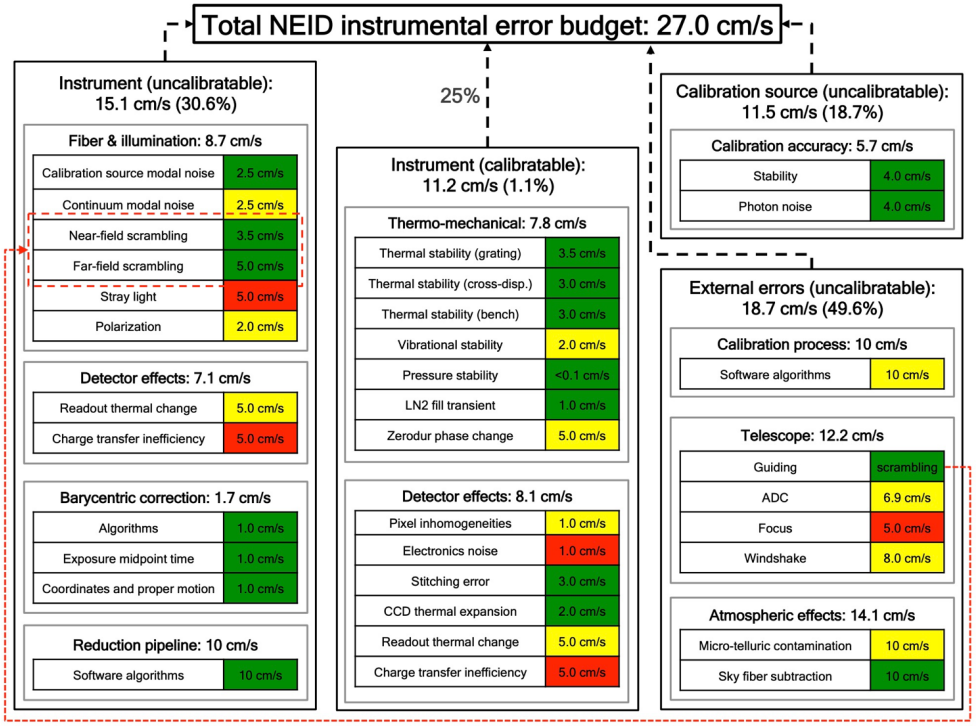


Figure 4.5: The RV error budget for the NEID instrument highlighting the significant contribution of an uncharacterised detector to the total error a measurement, (Halverson et al. 2016).

the following sections: fiber and illumination, detector effects (uncalibratable), barycentric correction, reduction pipeline, thermo-mechanical, detector effects (calibratable), calibration source, calibration process, telescope, and atmospheric effects.

Relevant to this work are the estimates of the calibratable and uncalibratable errors arising from geometric detector defects from both stitching boundaries and random pixel positions. The authors estimate that the calibratable detector effects (stitching boundaries, electronic noise, thermal expansion) comprise $0.11\,\mathrm{m\,s^{-1}}$, and that the uncalibratable detector effects (readout thermal change, charge transfer efficiency) comprise a further $0.07\,\mathrm{m\,s^{-1}}$ to the total budget. These values are added in quadrature with the other components to obtain the final total error budget.

This work highlights the estimated significant contribution of the detector to the overall error of a single measurement. Should these parameters be left uncharacterised, the error on each measurement would be an order of magnitude higher, as seen in Dumusque et al. (2015b). The full error budget estimation is shown in Figure 4.5.

4.4 The Pixel Mapping Procedure

To measure the geometric positions of an optical detector's pixels, we require some kind of optical 'ruler' whose regularity acts an an absolute reference. Interference fringes from coherent sources of light, such as lasers, provide an excellent example of this optical reference as the interference pattern created can be analytically described and modelled, and the regularity far exceeds the measurement precision required.

4.4.1 Modelling Laser Interference with Fourier Optics

To analytically describe the far-field intensity pattern from two interfering source I follow the procedure of Voelz (2011). This tutorial text guides the user through Fourier theory and eventually into far-field Fraunhofer diffraction expressions for various apertures and provides the user the *tools* to create their own diffraction patterns. The method is computationally efficient as it involves breaking down the far-field pattern into a discreet two-dimensional grid for evaluation. This is a perfect set of tools to begin simulating a CCD that is imaging a combination of apertures. This textbook is a MATLAB tutorial, for my purposes I converted the code into PYTHON. The particular form of the interference pattern arose from a private discussion of this specific implementation with the author (private communication, Voelz, 2018). In this subsection I give a brief description of the Fourier optics used for this simulation and how it gives rise to the 2D interference pattern as seen by the CCD. For the interested reader I recommend the aforementioned textbook as a starter for Computational Fourier Optics.

An optical source usually involves a two-dimensional function $g(x, y)$ to describe an aperture, such as a circular aperture. The analytic Fourier transform of such a function is given by

$$G(f_x, f_y) = \iint_{-\infty}^{\infty} g(x, y) \exp\big[- j2\pi(f_x x + f_y y)\big] dx dy \tag{4.1}$$

where $G(f_x, f_y)$ is the result of the transform and f_x, f_y are independent frequency variables associated with the two dimensions x and y. This can be noted in shorthand as $\mathfrak{I}\big\{g(x, y)\big\} = G(f_x, f_y)$. The inverse of the Fourier transform is given by

$$g(x, y) = \iint_{-\infty}^{\infty} G(f_x, f_y) \exp\big[j2\pi(f_x x + f_y y)\big] df_x df_y. \tag{4.2}$$

The shorthand for this operation is given by $\mathfrak{I}^{-1}\big\{G(f_x, f_y)\big\} = g(x, y)$. For an arbitrary source plane of monochromatic light propagating, the illumination aperture $U_1(\xi, \eta)$ is related to the observer plane $U_2(x, y)$ by the Rayleigh-Sommerfeld relation of

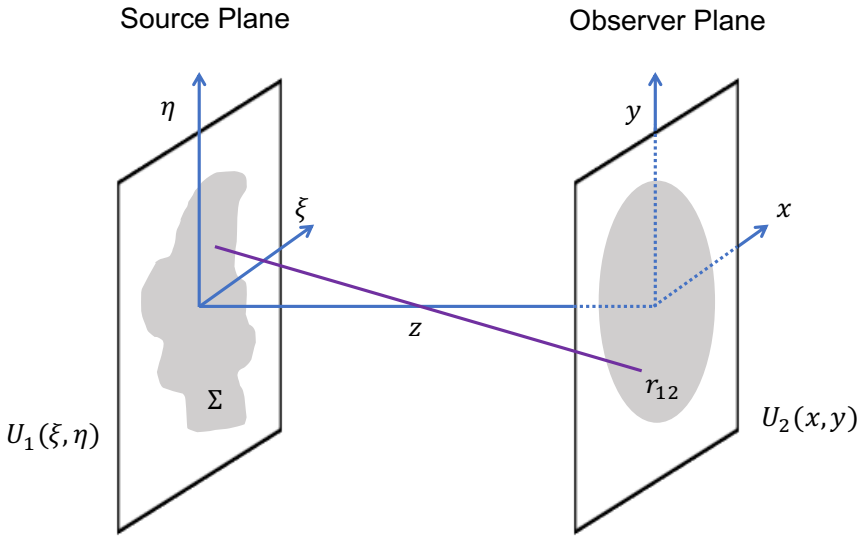


Figure 4.6: The coordinate system used to convert between the source plane and the observer plane.

$$U_2(x,y) = \frac{z}{i\lambda}\iint_{\Sigma} U_1(\xi,\eta)\frac{\exp(ikr_{12})}{r_{12}^2}d\xi d\eta, \tag{4.3}$$

where λ is the source wavelength, i is the imaginary number, k is the wavevector, z is the distance between the two plane centres, $r_{1,2}$ is the distance between a point on a source plane and a point on the observer plane, ξ and η are the variables of the integration enclosed by area Σ. This is graphically demonstrated in Figure 4.6. Because the source plane and the observer plane are defined to be parallel to each other, Equation 4.3 becomes a convolution integral of the form

$$U_2(x,y) = \iint_{\Sigma} U_1(\xi,\eta)h(x-\xi,y-\eta)d\xi d\eta, \tag{4.4}$$

where h is the impulse response of the system and is given by

$$h(x,y) = \frac{z}{i\lambda}\frac{\exp(ikr)}{r^2} \tag{4.5}$$

and $r = \sqrt{(z^2+x^2+y^2)}$. A convolution integral is usually very computationally expensive. But, it can be solved via a Fourier transform from the Fourier convolution theorem. Equation 4.4 becomes

$$U_2(x,y) = \Im^{-1}\{\Im\{U_1(x,y)\}\Im\{h(x,y)\}\}. \tag{4.6}$$

In this regime the integration variables between the two planes can be simply relabelled as x and y to give

$$U_2(x,y) = \Im^{-1}\{\Im\{U_1(x,y)\}H(f_x,f_y)\}, \tag{4.7}$$

where H is the Rayleigh-Sommerfeld transfer function and is given by

$$H(f_x, f_y) = \exp\left(ikz\sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2}\right) \tag{4.8}$$

With this formulation we only require that $r >> \lambda$ which is trivial as our wavelength is 400 nm to 700 nm and the propagation distance is of order 1 m.

We then invoke the Fresnel approximation to remove the square-root distance term, r, in Equation 4.3 as these make analytic expressions difficult to compute. A binomial expansion of

$$\sqrt{1+b} = 1 + \frac{1}{2}b - \frac{1}{8}b^2 ..., \tag{4.9}$$

where $b < 1$ allows us to expand r_{12} to

$$r_{12} \approx \left[1 + \left(\frac{1}{2}\frac{x-\xi}{z}\right)^2 + \left(\frac{1}{2}\frac{y-\eta}{z}\right)^2\right]. \tag{4.10}$$

When this approximation is applied to Equation 4.3, we obtain the Fresnel Diffraction Equation:

$$U_2(x,y) = \frac{e^{ikz}}{i\lambda z}\iint U_1(\xi,\eta)\exp\left\{i\frac{k}{2z}\left[(x-\xi)^2+(y-\eta)^2\right]\right\}d\xi d\eta. \tag{4.11}$$

This expression is also a convolution of the form in Equation 4.4, so the impulse response and transfer function can be written as

$$h(x,y) = \frac{e^{ikz}}{i\lambda z}\exp\left[\frac{ik}{2z}(x^2+y^2)\right], \tag{4.12}$$

and

$$H(f_z, f_y) = e^{ikz}\exp\left[-i\pi\lambda z(f_X^2, f_Y^2)\right]. \tag{4.13}$$

The expressions of Equations 4.6 and 4.7 become applicable here for computing the diffraction results. We can also rearrange the Fresnel diffraction equation by moving the quadratic phase terms of x and y outside of the integral to obtain

$$\begin{aligned} U_2(x,y) =& \frac{\exp(ikz)}{j\lambda z} exp\left[i\frac{k}{2z}(x^2+y^2)\right] \\ &\times \iint\left\{U_1(\xi,\eta)\exp\left[i\frac{k}{2z}(\xi^2+\eta^2)\right]\right\}\exp\left[-i\frac{2\pi}{\lambda z}(x\xi+y\eta)\right]d\xi d\eta. \end{aligned} \tag{4.14}$$

Fraunhofer diffraction refers to the 'far-field' regime where the following assumption is made

$$z >> \left(\frac{k(\xi^2+\eta^2)}{2}\right)_{\max}, \tag{4.15}$$

which results in the Fraunhofer diffraction equation[d]:

[d]The Fraunhofer approximation can also be considered if $z >> \frac{w^2}{\lambda}$, which for our purposes is easily satisfied.

$$U_2(x,y) = \frac{\exp(ikz)}{j\lambda z} exp\left[i\frac{k}{2z}(x^2+y^2)\right] \times \iint \left\{U_1(\xi,\eta)\exp\left[-i\frac{2\pi}{\lambda z}(x\xi+y\eta)\right]\right\} d\xi d\eta. \tag{4.16}$$

The Franhofer diffraction equation can be recognised as a Fourier transform of the source field. Hence, the far-field irradiation of a simple aperture such as a circle or square, can be analytically calculated. For a simple circular aperture illuminated by a plane wave of unit amplitude, the complex field immediately beyond the aperture is given by[e]

$$U_1(\xi,\eta) = \text{circ}\left(\frac{\sqrt{\xi^2+\eta^2}}{w}\right). \tag{4.17}$$

The Fourier transform of a circle is a Bessel function of the first kind order 0 - a sinc function. Continuing to follow the notation of Voelz (2011), this is given by

$$\Im\{U_1(\xi,\eta)\} = w^2 \frac{J_1\left(2\pi w\sqrt{f_\xi^2+f_\eta^2}\right)}{w\sqrt{f_\xi^2+f_\eta^2}}, \tag{4.18}$$

where J_1 indicates the Bessel function and w is the radius of the circular aperture. The frequency variables can be substituted with $f_\xi = x/\lambda z$ and $f_\eta = x/\lambda z$, and then Equation 4.16 becomes

$$U_2(x,y) = \frac{\exp(ikz)}{i\lambda z} exp\left[i\frac{k}{2z}(x^2+y^2)\right] \times w^2 \frac{J_1\left(2\pi\frac{w}{\lambda z}\sqrt{x^2+y^2}\right)}{\frac{w}{\lambda z}\sqrt{x^2+y^2}}. \tag{4.19}$$

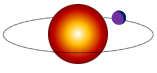
Optical detectors do not yet exist that can sample the very high frequency oscillations of the electromagnetic fields of $\sim 10^{14}$Hz. Instead, they respond to the time-averaged squared magnitude of the field. Hence, the *irradiance* (W m^{-2}) is the quantity of interest:

$$I(x,y) = \left[U_2(x,y)\right]^2 = \left(\frac{w^2}{\lambda z}\right)^2 \left[\frac{J_1\left(2\pi\frac{w}{\lambda z}\sqrt{x^2+y^2}\right)}{\frac{w}{\lambda z}\sqrt{x^2+y^2}}\right]^2. \tag{4.20}$$

This function is plotted in the left half of Figure 4.7. The Bessel function is a rotationally symmetric sinc function, and the x and y parameters are the coordinates of the sampling grid used.

I model the optical fibers as a pair of circular apertures of equal radius and on the same optical plane as described above. Here, the source plane is given by the sum of two circular apertures similar to Equation 4.17 as (see Figure 4.8):

$$U_1(\xi,\eta) = \text{circ}\left(\frac{\sqrt{(\xi-\xi_1)^2+(\eta-\eta_1)^2}}{w}\right) + \text{circ}\left(\frac{\sqrt{(\xi-\xi_2)^2+(\eta-\eta_2)^2}}{w}\right). \tag{4.21}$$



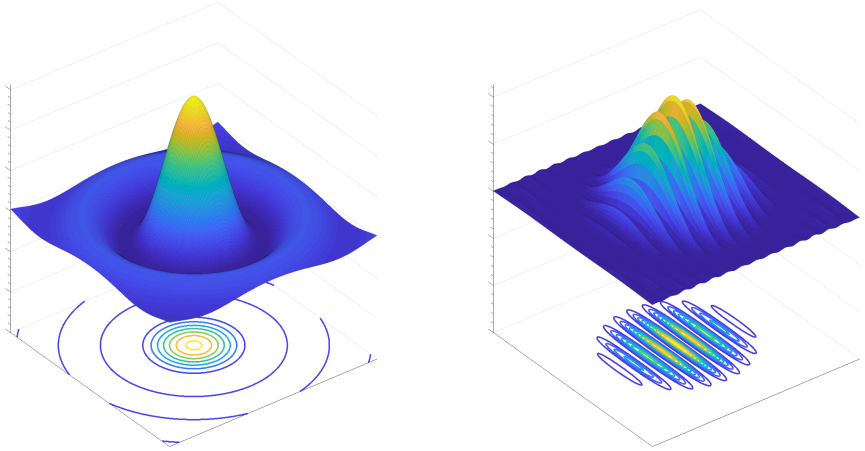


Figure 4.7: The 2D Fourier transform of a perfect circular aperture, left, results in a rotationally symmetrical sinc function. This is the far-field illumination pattern that a single optical fiber produces. When a second source is introduced, right, the intensity pattern is the convolution of the two functions evaluated by a grid of points in the far-field plane. Tuning the parameters of the two fiber sizes and orientations allows for any physically possible interference pattern to be simulated.

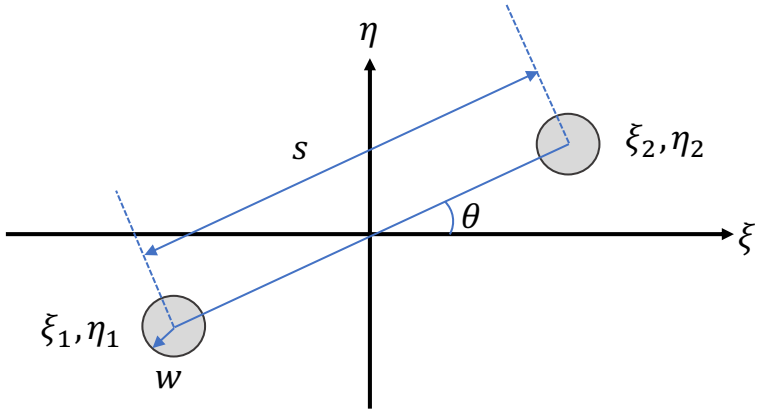


Figure 4.8: The arrangement of the two fibers in the optical plane annotated with the relevant parameters needed for Fraunhofer diffraction.

Following the previous procedure but with Equation 4.21 as the source plane aperture will yield the far-field Fraunhofer diffraction pattern from two adjacent circular apertures. A few substitutions can be made to simplify the process, as seen from Figure 4.8 the spatial integration variables can be written as $\xi = \frac{s}{2}\sin(\theta)$ and $\eta = \frac{s}{2}\cos(\theta)$. The sampling grid can be thought of

[e] 'Circ' refers to a cylindrical top hat function that equals 1 inside of $\sqrt{\xi^2 + \eta^2}$, and zero otherwise.

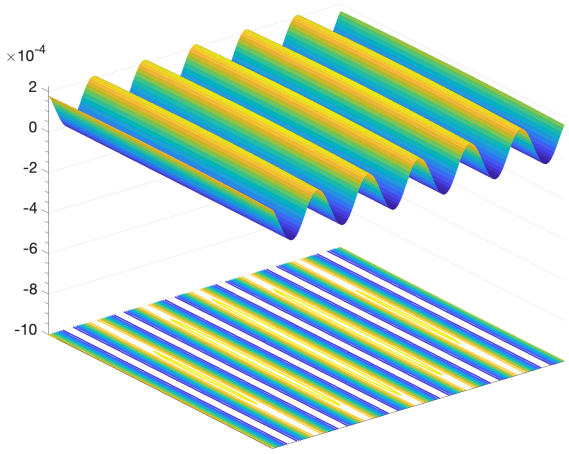


Figure 4.9: Optical fringes produced from Equation 4.22. The parameters of the geometry of the optics are tuned such that useful fringes are produced. The z-scale highlights the small fraction of power that is collected by the pixels in this regime.

X and Y coordinates of pixels used for sampling the far-field pattern.We then explicitly write the Bessel function as a sinc function and tidy up some of the physical parameters to obtain

$$I_{(i,j)} = \left(\frac{4w^2}{\lambda z}\right)^2 \text{sinc}^2\left(\frac{w}{\lambda z}(X^2+Y^2)^{1/2}\right)\cos^2\left(\frac{\pi s}{\lambda z}(X\cos\theta + Y\sin\theta) + \phi\right). \tag{4.22}$$

Here, λ is the wavelength of light used, z is the propagation distance between fibers and detector, w is the diameter of the fiber core, X and Y are the coordinates of a given pixel (i, j), s is the magnitude of separation between two fibers, and θ is the angle between the two fibers from horizontal. Figure 4.7 shows how the irradiance pattern is modulated with the presence of a second circular aperture.

When the optical parameters are augmented such that the aperture radius is of order µm and their separation is of order mm, a distant detector of small pixels sees almost uniform fringes. The sinc envelope that is shown in Figure 4.7 is now much larger than the detector to the effect that the overall intensity profile is nearly flat. Such a configuration is shown in Figure 4.9.

Figure 4.10 shows a schematic of the experimental setup and how the physical optical parameters relate to those mentioned in the Fraunhofer diffraction. With this formulation, each pixel will have a value between 0 to 1 where the value is the fraction of the total power emitted by the lasers collected by that pixel. Because the pixels are very small relative to the total field of light emitted, and there are upwards of 4 million pixels for a 2048×2048 sized detector, these numbers are typically of the order 10^{-11} at maximum. It is sensible to linearly scale the image such that the maximum value is 1, or convert these intensity values into actual counts, which is

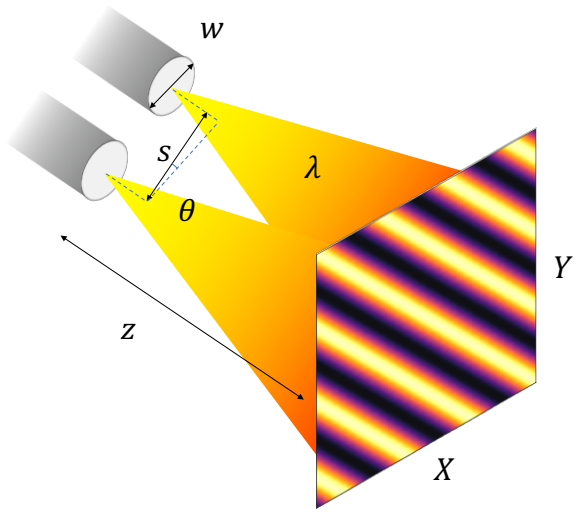


Figure 4.10: A schematic of two fibers of aperture diameter w, separated by s at angle θ to horizontal, propagating laser light of wavelength λ across a distance z to a detector of pixel coordinates X_i, Y_j in a plane perpendicular to the fiber apertures. Where the two beams overlap they interfere, and the CCD images this interference pattern.

the general case for CCDs.

The actual number of counts, $C_{i,j}$ registered by each pixel is computed by

$$C_{(i,j)} = I_{(i,j)} \cdot \frac{\lambda}{hc} \cdot Q \cdot t_{exp} \cdot P_{laser} \cdot L, \tag{4.23}$$

where h is Planck's Constant, c is the speed of light, Q is the QE of the detector at that wavelength, t_{exp} is the exposure time, P_{laser} is the power of the laser, and L is the fractional losses of the optical system from laser emission to fiber tips. In a simulation, this value then gets rounded down to the nearest integer. Further constraints could be added to this conversion to factor in the gain of the CCD, and limit the counts at some saturation level.

I use Equation 4.22 to determine the optimum geometric configuration such that the optical fringes are most useful for a practical experiment. I know that the spectral range of the HARPS3 CCD lies between 380 nm to 680 nm which constrains the choice of laser wavelength. I also know that the optical bench I will use is 3× 1.5 m which constrains the propagation distance and fiber spacing baselines. Lastly, I anticipated that useful fringe periodicity lies between 10 and 100 pixels. These bounds arise from Nyquist sampling considerations, and ensuring that the 2D spatial fit has enough structure to well confine the parameters. Lastly, I predicted I would chose single-mode fibers to remove the speckle pattern often seen with multi-mode fibers. This limits the aperture diameter to around 5 µm. I discuss these constraints further in Section 6.2.

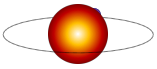
4.5 Next Steps

Towards the efforts of an end-to-end pixel mapping experiment and data analysis pipeline, I conducted three main projects aimed at achieving this result. I first created a basic interferometer to investigate the disturbances to the delicate interference pattern that comes from the environment in the form of vibrations and thermal drift from our lab.

Secondly, I then created an end-to-end pipeline that models a CCD, gives each pixel a displacement drawn from a distribution, and then images an interference pattern. The interference pattern is a full 3D model of the far-field Fourier optics of two interfering laser beams, Equation 4.22, and the intensity field is sampled by each pixel. The pipeline then analyses the CCD images and extracts the injected displacements. I then test this pipeline in a variety of different circumstances such as varying the magnitudes of the displacements, varying the injected photon noise, and assessing the limit of the required number of measurements.

Thirdly, I use this simulation to design an all-fiber optical setup to generate the fringes as required by the analysis. I built and tested this optical setup and used it to measure the pixel positions of two optical detectors, one of which shares the same pixel architecture of the actual HARPS3 science CCD.

In the next chapters I discuss these three projects and the early results from this experiment. I also include many suggested improvements, next steps, and complimentary work that could be conducted to further refine and develop this technique. Ultimately, the aim is to produce a 'black-box' which could be used on any optical detector to capture the relevant data, and then process it to produce a pixel-displacement map.



Chapter 5

A Simulation and Data Analysis Pipeline to Measure the Pixel Positions of a CCD

> 'All we ever see of stars are their old photographs.'
>
> *Dr. Manhattan*

In Chapters 1 and 4 I introduced the concept that the positions of the pixels of a detector may not be a perfect grid in Cartesian coordinates. Whether from imperfections in the fabrication processes, or warping from temperature variations and mechanical stress, a detector's pixels may deviate off this assumed grid.

For everyday applications of optical detectors such as amateur photography this deviation is negligible. But rigorous scientific applications, such as spectroscopy, may use the pixels as measuring tool themselves. In the case of high-precision radial velocities, the pixels are used as a ruler to measure the drift of spectral lines, which is then inferred as the presence of planets. Large planets cause the spectra to drift across a significant fraction of a pixel and in these cases the pixel position deviations are too small to notice. But for Earth-twin discoveries on an instrument such as HARPS3, the magnitude of the spectra motion may be as little as $1/1000$ of a pixel, or around 15 nm. In these cases, if the pixel positions have an uncertainty of around 0.1%, then the error on the Earth-twin could be as much as 100%.

The aims of this project are to create an experiment and analysis pipeline that has the capability of measuring the pixel positions of a CCD to $1/1000$ of a pixel. This ambitious project has a few major subgroups. Firstly, the data analysis needs to be robust to handle

various optical interference patterns from CCDs of different sizes, but needs to scale well with larger detectors and datasets. Secondly, the experiment needs to be physically plausible. The analysis pipeline was created in such a way that it can be used in 'simulation mode' to practice the analysis on simulated data. Here, the user can trial different optical components and setups to see the effect on analysis. Thirdly, the environment effects need to be considered. Interference fringes are incredible sensitive to small changes in temperature and vibrations, the experiment design needs to mitigate this.

5.1 Optical Simulation and Analysis Procedure

The technique described in this section is largely based on a similar experiment conducted by Shaklan et al. (1995), and later by Crouzier et al. (2012). In these experiments, the authors created their interference pattern with two laser beams, but modelled the interference pattern as a simple sine wave as compared to the full Fraunhofer model I discussed in Section 4.4.1. Whilst vastly increasing the speed of computation, this assumption falls short for pixels that lie towards the edge of a large detector. The fringes generated by the simulation (and those that will be captured by the experimental setup) are *not* uniform in spatial frequency. This is because the intensity pattern emitted is in spherical wavefronts whilst the CCD images a 2D cross-section of this field. Hence, if an analysis script simply takes a 2D FFT of the data and uses this to generate a 'fitted' image, the fringes towards the edge of the detector will not be described appropriately. The technique used by Shaklan et al. (1995) did make an adjustment to their model to account for this by including a hyperbolic term as a function of distance for a pixel away from the centre of the detector. However, their model did not also account for the overall sinc envelope that covers the optical fringes. The implementation of this technique from Crouzier et al. (2012) was for a much smaller format CCD than that of HARPS3. Theirs was just 80×80 pixels which meant the overall sinc envelop was almost negligible for their detector. Plus this detector could be read out at speeds many times faster than the predicted read-out time of HARPS3 which greatly reduced the need for short term fringe stabilisation plus this CCD had zero stitching boundaries due to its size.

Another technique of measuring pixel positions involves using a flat-field image from an LED to measure the block-stitching boundary size to a precision of ±0.005 pixel, which is then verified with a Laser Frequency Comb (Coffinet et al. 2019). However, the authors note that this technique is not sensitive to the pixel displacements within a block which is the aim of this work. But, due to their precision, LFCs can also be used to measure the intra-pixel variability, i.e. how the photometric sensitivity of the pixel varies across its surface (Ravi et al. 2017). These measurements provide insight into tracking and monitoring the reference spectra from

an LFC to well below 1 m s^{-1}, but do not provide information on the actual coordinates of the pixels themselves. An LFC, by design, only illuminates a small fraction of the pixels at any given time as it only projects a series of discreet peaks. Hence, using one for pixel position measurement is not suitable for full CCD characterisation whereas fringes *are* imaged by the whole CCD at every frame taken.

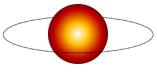
To aid in designing the final experiment, and to better understand the results of the analysis procedure, the optical simulation is conducted in a few steps:

1) Define a set of optical parameters:
   a) Laser power and wavelength
   b) Geometry of optics
   c) CCD pixel size and number (M_x, M_y)
2) Create an x, y displacement for each pixel from a distribution
3) Using displaced pixels, sample the 2D interference pattern from Equation 4.22
4) Compute a 2D spatial fit of the data from the same equation
5) Sweep the phase N times, and repeat $1-4$ to create 2 data cubes of size $[M_x, M_y, N_{\text{phase}}]$
6) For each pixel compute:
   a) 1D sine-wave temporal fit to a pixel of the real data
   b) 1D sine-wave temporal fit to the same pixel from the spatial fit
   c) Calculate the phase difference
   d) Compute the pixel displacement along the fringe wavevector
7) Compare the measured results to the injected displacements
8) Add white noise and repeat $1-6$
9) Modify the injected displacements and repeat $1-6$
10) Add stitching boundaries and repeat $1-6$

A natural by-product of this procedure is the data-analysis scripts for the actual experimental data. Once the simulation is conducted with a certain level of 'realism' and the user is happy it is ready to accept experimental data, only a few lines of code need to be changed to point the scripts to the actual data files.

5.1.1 Choosing Optical Parameters

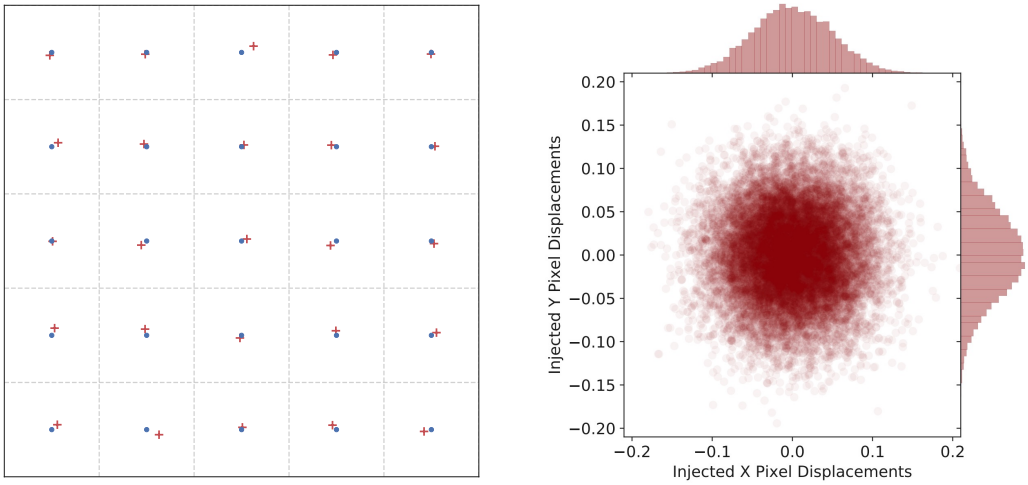
For item 1 in the procedure, the optical parameters can be chosen with some understanding of Equation 4.22. It is easy to see that, for example, increasing the propagation distance z results in longer period fringes simply due to the inverse-square law of propagation but also results in dimmer fringes. Narrowing the fiber-core w will broaden the *sinc* envelope of the overall pattern. Moving the fibers closer with s will increase the periodicity, and changing their relative angle θ will rotate the fringes on the plane of the detector.



Many of these parameters are already constrained due to what components are available at a sensible price and the physical constraints of the lab. The propagation distance, z, is restricted to the longest length of the optics bench minus some space either side for miscellaneous components and circuitry, hence 1.5 m was nominally chosen as the minimum propagation distance which gave room for more should I need it. It was known the fibers would be single-mode (SM) to mitigate the speckle pattern observed when using a multi-mode (MM) fiber. SM fibers have very narrow cores of just a few µm which makes optical alignment difficult, but massively increases the FWHM of the central lobe of the *sinc* function. A CCD at $z = 1.5$ m illuminated by a $w = 5$ µm core only sees ~1% fringe brightness variation from edge to edge (the full-size CCD is just over 6 mm wide) as the *sinc* function central lobe is very broad here. The physical width of the fiber and cladding limits the minimum spacing, s, they can be placed together. Standard 'off-the-shelf' SM fibers can be ordered with a 2.5 mm ceramic ferrule tip for protection of the fiber core, hence 2.5 mm is the minimum separation for a protected pair of fibers. Should this not be sufficient, i.e. the simulation dictates the fibers need to be closer together for broader fringes, the ceramic tip can be removed to reveal the delicate fiber core and cladding of just 0.125 mm in diameter. As these are a custom order, the first iteration of simulations was conducted assuming the ceramic tips are in place to see if it was sufficient. Lastly, the angle between the fibers θ does not affect the fringes apart from their rotation relative to the detector. One could also rotate the detector to achieve the same result. As I will show in the next section, the fringes probe *both* the x and y pixel displacement vector when the fringes are $0° < \theta < 90°$, i.e. *not* exactly perpendicular or parallel to the CCD axis, so I choose 45° as a starting point. The exact components used will be listed after the preliminary results of the simulation.

5.1.2 Simulating a CCD with Pixel Displacements

The CCD is modelled as a 2D grid of points pitched at the nominal pixel size. Each point on the grid represents the coordinates of the *centre* of that pixel, and the whole CCD is centred at the origin of the optical axis. For each pixel, a unique δx and δy is drawn from a normal distribution of $\mu = 0$ with some σ determined by the user. Figure 5.1a shows a section of a simulated CCD with the pixel centres deviating from their assumed position. For exaggeration in this plot the 1σ width is 0.05 pixels, this bivariate distribution is shown in Figure 5.1b. For simplicity in the following sections, the images and data shown are from a 128×128 pixel detector.



(a) Pixel centres as they appear on the CCD versus their true hidden positions.

(b) All injected displacements.

Figure 5.1: a) 5×5 pixel zoom of a simulated detector with pixel displacements and b) the distributions of all injected pixel centres for a simulated 128×128 pixel detector, units are in pixels. For a) the blue dots represent the assumed centres of each pixel, the red crosses are the actual injected displacements.

5.1.3 Sampling the Interference Pattern

With the x and y displacements added to each of the pixel coordinates, a simulated image of the far-field interference pattern from two optical sources can be generated. In Equation 4.22, copied below for reference, the sampling grid coordinates X and Y are replaced with the new pixel coordinates and the other optical parameters are taken from Section 5.1.1.

$$I_{(i,j)} = \left(\frac{4w^2}{\lambda z}\right)^2 \operatorname{sinc}^2\left(\frac{w}{\lambda z}(X^2+Y^2)^{1/2}\right)\cos^2\left(\frac{\pi s}{\lambda z}(X\cos\theta + Y\sin\theta)+\phi\right). \tag{5.1}$$

For this simulated detector, the fringes are shown in Figure 5.2. These are 'perfect' fringes in that the detector has zero pixel displacements, a perfect flat-field correction, zero dark current, and zero photon limited shot noise. The pixel displacements, flat-field, and photon noise are included in the simulation. The parameters which define these noise sources can be set to any value desired *before* the simulated CCD images the interference pattern to ensure that each data-set has a consistent set of noise sources.

5.1.4 The 2D Spatial Fit

For a detector that has no stitching boundaries but millions of pixels, the pixel displacements can be approximated with a normal distribution of $\mu = 0$, i.e. the pixel displacements are randomly distributed. In this regime, the spatial 2D fit can be a least-squares minimalisation

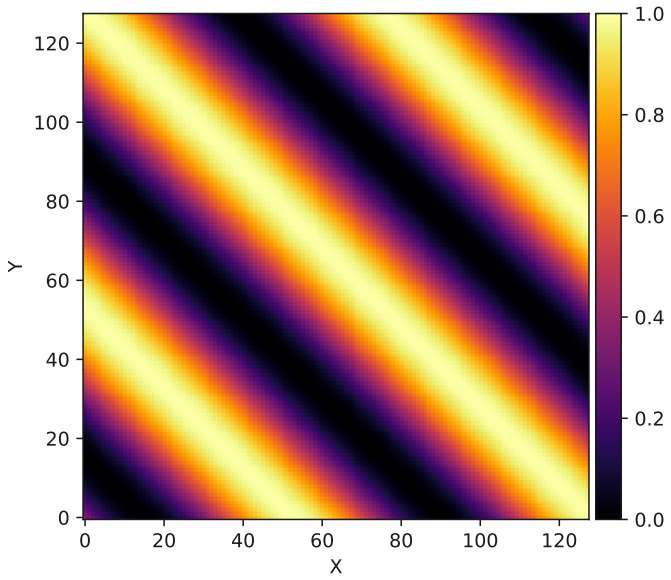


Figure 5.2: The optical interference pattern generated from Equation 5.1 with the parameters as detailed in Section 5.1.1. The image has been normalised between 0 to 1. This particular optical set up yields fringes at a period of around 50 pixels in the wavevector direction.

of the interference model of Equation 4.22. The fitted parameters can then be used to generate an image of what the CCD would register if it were perfect and defect free. This fit is repeated at each phase step to build the data cubes. Hence, the temporal response of each pixel in the fitted data cube is what they *should* be registering. If the stitching boundaries are known, the CCD can simply be analysed in blocks of that size and the results combined to create a full map of the detector. Ultimately, these measured positions will be used in the data reduction pipeline of HARPS3 as a correction step to convert the measured positions of the spectra into the true positions of the spectra and used in the calculation of the wavelength solution - where the position of the spectral calibration lines are accurately mapped to the science spectrum.

The data from Figure 5.2 is then subject to a spatial fit[a]. The function fitted has the form of Equation 5.1 where X and Y are the assumed coordinates of the detector, i.e. the model will generate fringes that are 'perfect' in structure. Because the fit has 6 parameters (w, λ, z, s, θ, ϕ), the fit is computationally expensive and contains degeneracies. For example, doubling the propagation distance z will result in fringes of half the period whilst doubling the aperture separation s will have the same effect.

To combat this the parameters are given tight bounds that are derived from the predicted physical uncertainties of each of the quantities. These bounds are estimates in the simulation

[a]The checker-board pattern that may be visible is a display artefact.

Table 5.1: Parameter examples for the 2D Spatial Fit used in the simulation. The values and their limits are updated when the experiment is conducted to represent the measurement uncertainty of each quantity. The phase cannot be known *a priori*, so is randomly chosen at each fit iteration.

Parameter	Symbol	Value	Lower Limit	Upper Limit
Propagation Distance	z	1.5 m	1.49 m	1.51 m
Fiber Separation	s	2.5 mm	2.4 mm	2.6 mm
Angle	θ	45°	44.5°	46.5°
Fiber Core Diameter	w	5 µm	4 µm	6 µm
Laser Wavelength	λ	532 nm	531.5 nm	532.5 nm
Phase	ϕ	-	0	2π

but can be updated when the experiment is created and are summarised in Table 5.1. It is anticipated that the physical distances and angle (z, s, θ) will be constrained to even further levels than that shown in the table as the fibers will be mounted in a custom-machined metal block with µm level precision. The laser wavelength is not anticipated to vary between frames, but it could vary slightly from the specification so it will be a variable for the first iteration and then fixed. The fiber-core diameter will also be further constrained after the first fit iteration.

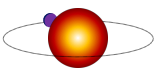
The fit is implemented with the PYTHON package SCIPY, and its library CURVE_FIT [b]. This fitting routine takes in some data and a function with a list of the parameters and their bounds, and returns the optimised least-squares fit of that function to the data in the form of a list of optimised parameter values. The simulation then takes these values and regenerates the interference pattern with a grid of non-displaced pixels.

This process results in a 'real' image of fringes, and the ideal image that would be produced given no photon-noise and no pixel displacements.

5.1.5 Sweeping the Phase

Then, with all other parameters now fixed, the simulation will regenerate the interference pattern with a slightly higher phase value and repeat the fitting process until 2π total phase shift has been completed. This has the effect of sweeping the fringes across the detector parallel with the wavevector. In the experiment it is anticipated that the fringe sweeping will be controlled via a delay line, phase modulator, or fiber stretcher. However, thermal expansions and contractions could inhibit the idealised uniform phase steps. Hence, the phase sweep in the simulation is given some slight random offsets which results in a non-uniform distribution of data in the temporal domain. This means that sufficient data needs to be collected to ensure the temporal

[b]The documentation on the fitting algorithm provided by the SCIPY library is found at: https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.curve_fit.html.



fit to each pixels response needs to be has enough phase-coverage to ensure good parameters. A starting point for the simulation is 100 phase-steps between $0 - 2\pi$ but can be set arbitrarily high.

At each phase-step, the 2D fit is computed and the data, spatial fit, and parameters stored in a data cube for archiving. The data cube is shown in Figure 5.3a. It is expected that for the experiment, the phase may not behave linearly or even sequentially. Thermal drifts between the two fiber channels may cause additional phase shifts that are uncontrollable but not to the extent where the data is useless. Hence, the 2 data cubes (data and the fit) are sorted as per the phase value of each 2D fitted slice such that the entire cube only covers one complete cycle with $0 < \phi < 2\pi$.

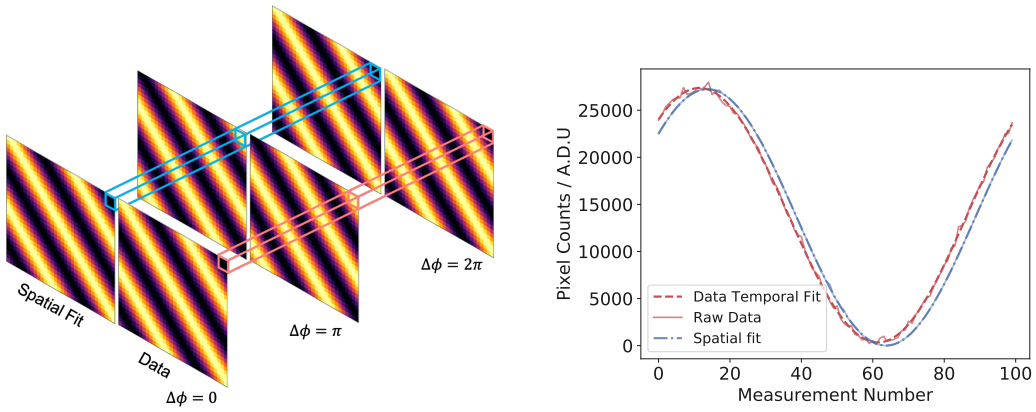
5.1.6 The 1D Temporal Fit

With the phase sweep complete, the simulation outputs 3 data cubes. The first is the real fringe data generated by the simulation, the second is an identical sized array of the 2D spatial fit to each of the frames, and the third is a list of fitted parameters at each slice. It is worth noting here that the data is large, it can have dimensions 2×2048×2048×100. For a 16-bit detector that stores 65536 levels of pixel counts, this array reaches 1.6Gb. When running the simulation multiple times with different levels of phase-sweep steps, computer memory becomes an issue. The simulation will write many of the variables to file where possible but not excessively as disk writing can be time consuming.

The data is then passed through a 1D fitting routine which looks at each pixels response as a function of time (phase). This process is graphically demonstrated in Figure 5.3b. Each of the 1D data arrays is passed through a sine-wave fitting function which returns the period, phase, amplitude, and offset of the curve. The relative phase values are of interest for pixel displacement calculations. This 1D fit is computed twice for each pixel, one for the real data and the other for the spatial fit. The temporal fit is fast as the bounds on the variables can be tightly constrained as when the data-cube is normalised, the period and amplitude are exactly 1, and the offset can be guessed from the lowest value in the data-cube.

5.2 Deprojecting the Measured Pixel Displacements

To obtain a 2D vector of coordinates from the temporal fits, two sets of perpendicular fringes from identical baselines need to be collected as shown in Figure 5.4. Because these two wavevectors are known to be orthogonal, the procedure for computing the pixel displacement δ_{ij} is as follows.



(a) A single data cube of optical fringes.

(b) 1D Temporal Fit

Figure 5.3: a) a data cube of fringes as a function of time (phase). The highlighted pixel's data is sent for temporal fitting. b) the phase difference between the spatial fit and the data of a given pixel highlighted in a). The 1D temporal signal from both curves is fitted with two, single period sine-waves whose phases are extracted for pixel displacement calculations. The phase difference between these two waves is proportional to the pixel displacement along the wavevector.

For each pixel, a sine-wave is fitted to the temporal response of a given pixel from the data and from the spatial fit (see Figure 5.3b), and this is conducted for each of the two baselines resulting in 4 fitted sine-waves per pixel. The sine-wave fit returns all the parameters that define a sine-wave but only the phase, ϕ_{ij}, is considered for pixel displacements.

Each iteration of the 1D temporal fit will compute a single pixel's displacement along the two perpendicular wavevectors by first calculating the phase difference between the 2D data and the 2D spatial fit:

$$\begin{aligned} \Delta\phi_{ij1} &= \phi_{\text{data\_}ij1} - \phi_{\text{fit\_}ij1} \\ \Delta\phi_{ij2} &= \phi_{\text{data\_}ij2} - \phi_{\text{fit\_}ij2}, \end{aligned} \tag{5.2}$$

where ij denotes pixel on row i column j, subscript 'data' is the 2D data from the fringes, subscript 'fit' is the 2D data from the spatial fit, and subscript '1' or '2' refers to fringe data-set of wavevector 1 or 2.

This phase difference is proportional to that pixel's displacement along the wavevector so it must be normalised to the wavelength of the 2D carrier wave to convert into absolute displacements in units of pixels with

$$\begin{aligned} \alpha_1 &= \Delta\phi_{ij1}/K_1 \\ \alpha_2 &= \Delta\phi_{ij2}/K_2 \end{aligned} \tag{5.3}$$

where $K_1 = 2\pi/\lambda_{2\text{D}}$ is the wavevector of fringe set 1 and $\lambda_{2\text{D}}$ is the 2D wavelength in pixels (e.g. $\sim$ 70 pixels as in Figure 5.5), and likewise for K_2. It is assumed that $\lambda_{2\text{D}}$ is identical

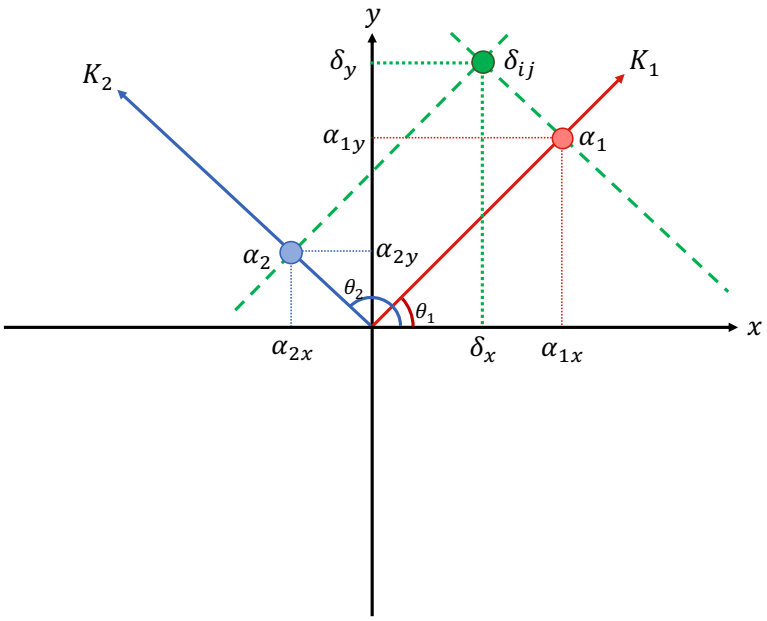


Figure 5.4: The geometry used to deproject the measured pixel displacements from two perpendicular baselines K_1 and K_2. The displacements measured are only along the wavevector of their respective fringes, but α_1 and α_2 can be perpendicularly deprojected (green dotted construction lines) to the true displacement $\delta_{i,j}$ of that pixel.

for both wavevectors as in practicality the block that holds the fibers will simply be rotated by 90°. Then, each measured displacement is written in orthogonal components which can then be independently summed for the final displacement

$$\begin{aligned} \alpha_{1x} &= \alpha_1 \times \cos(\theta_1) \\ \alpha_{1y} &= \alpha_1 \times \sin(\theta_1) \\ \alpha_{2x} &= \alpha_2 \times \cos(\theta_2) \\ \alpha_{2y} &= \alpha_2 \times \sin(\theta_2) \end{aligned} \tag{5.4}$$

where $\theta_{1,2}$ is the angle of the fringes for each of the wavevectors. These components are then summed in each of the x and y axes for the final 2D vector that describes the true pixel displacement

$$\begin{aligned} \delta_x &= \alpha_{1x} + \alpha_{2x} \\ \delta_y &= \alpha_{1y} + \alpha_{2y}. \end{aligned} \tag{5.5}$$

This procedure is computed *per pixel* once all of the 2D fits have been completed. Because these iterations are completely independent of any other pixel, they can be practically coded with a function that allows for multi-threaded processing. For my purposes, all of the results

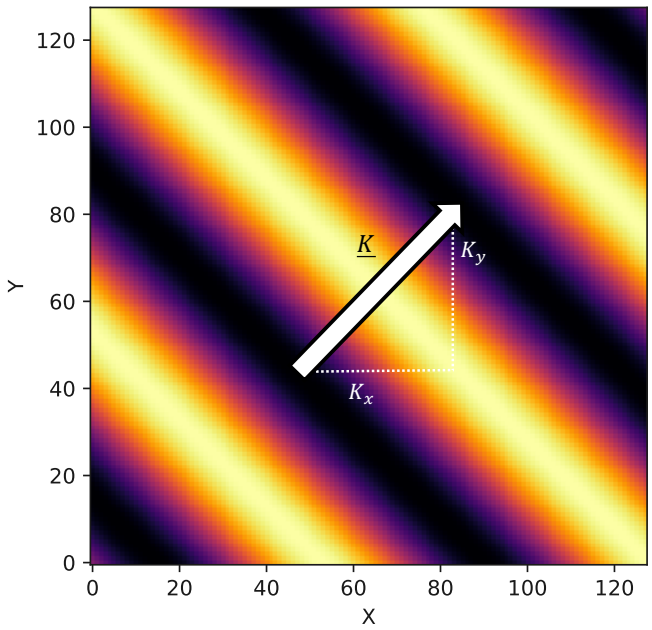


Figure 5.5: A 128 by 128 pixel subsection of a simulated interference pattern formed by two distant laser beams imaged by an ideal CCD. The wavevector $\boldsymbol{K}$ is perpendicular to the fringes and points to the apparent motion of the fringes as the phase is swept.

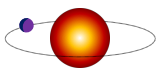
were conducted on the Cavendish Astrophysics Exoplanet cluster which had 128 available threads. This massively reduces the computation of pixel displacements for the largest detector considered as this procedure in total could fit upwards of 16 *million* sine-waves. Parallelising this reduces the computation time to the order of a few hours rather than a few months.

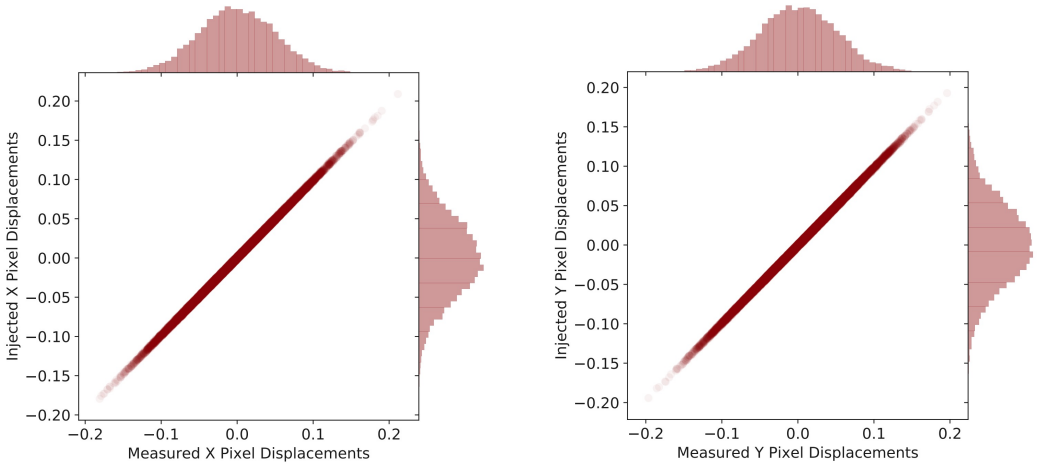
5.3 Simulation Results

In the following sections I detail the results from the end-to-end optical simulation from generating fringes through to pixel displacement measurements. In each subsection I sequentially increase the realism of the simulated data by including photon limited shot noise, larger CCD formats, and stitching boundaries. The results here are then used to guide the experiment design and component selection.

5.3.1 Simulation Results - No Noise Sources

In this section I use the aforementioned procedure of Sections 5.1.4 to 5.2 on a set of simulated CCDs that only contain normally distributed pixel displacements. Photon noise, stitching boundaries, and non-uniform flat fields are omitted. This procedure acts as a test of the analysis script in its ability to find pixel displacements in an ideal scenario.





(a) Injected and measured displacements in x

(b) Injected and measured displacements in y

Figure 5.6: a) the bivariate distribution for the measured and injected horizontal pixel displacements from a simulated CCD. b) the same for the vertical displacements.

For a starting point I use a CCD of size 128×128 pixels, the two fibers are 5 mm apart at 45°, the propagation distance is z =1.5 m, the fiber core diameter is 5 µm, and 100 phase steps are taken. The scale of pixel displacements here is a Gaussian distribution of $\mu = 0, \sigma = 0.05$ pixels. For the spatial fit, these parameters are given as bounds with 5% errors on each and no first-guess is given. I.e. I tell the spatial fit the range of the allowed parameter space to search for each parameter. It is anticipated that the actual uncertainty of these bounds will be greatly reduced for the real experiment.

For large numbers of pixels it is not practical to investigate each pixel in turn to check the accuracy and precision of the measurements. I plot the bivariate distributions of the injected and measured displacements in each axis on the CCD to investigate the correlation. Figure 5.6a and 5.6b show the very strong correlation for both of the measured δ_x and δ_y distributions. However, as a sanity check, I have re-plotted Figure 5.1a now with the measured displacements added also, see Figure 5.7, which shows that the measured displacements agree well with the injected displacements. There are some small slight deviations from the true displacements as seen by slight non-zero width of the bivariate distributions of Figure 5.6. If the analysis had measured each pixel perfectly, the bivariate distribution would show a straight line with zero width. To investigate this, I have plotted the distribution of the absolute differences between the measured and injected displacement 2D vectors in Figure 5.8.

The differences distribution shows that the measurements on the simulated CCD have good accuracy. At most, the difference between the measured displacements and the injected is 0.0042 pixels which is just over 4× the target precision of 1 milli-pixel. However, this is just from one baseline and with one measurement. It is expected that with each set of subsequent

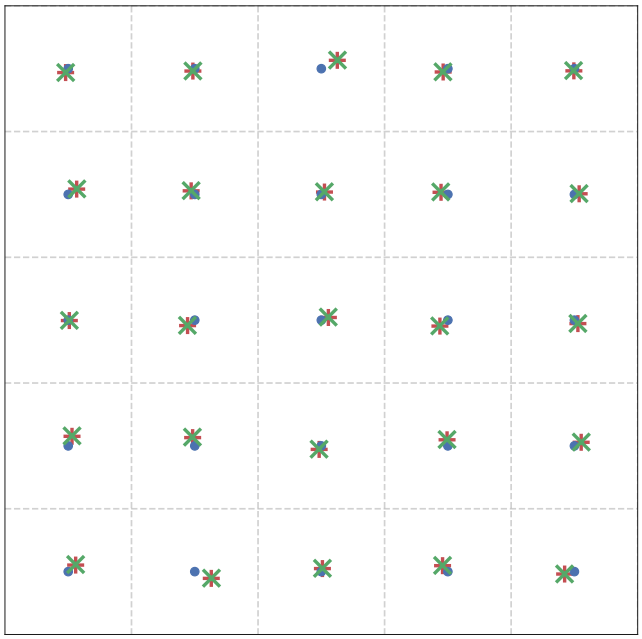


Figure 5.7: The measured pixel displacements as shown in Figure 5.1a. The assumed pixel centres are the blue circles, the injected pixel displacements are the red '+' symbols, and the measurements are green '×' markers.

measurements, the differences between the measured and real displacements will decrease. This is investigated further in Section 5.3.3. The distribution is also not centred exactly 0, the fitting algorithm has underestimated the Y displacements and overestimated the X displacements on average by approximately 1 milli-pixel. This could be a symptom of the algorithm in general or specific to this particular set of fringes.

The peak of this 2D distribution is at $x = -0.00033, y = -0.00024$ pixels, very close to zero, and the standard deviation for each set is $\sigma_x = 0.0013$, $\sigma_y = 0.0013$ pixels respectively. This result says that with zero photon noise, the pixel mapping procedure regularly obtains pixel measurements precisions to just over 1 milli-pixel 68% of the time. Ideally, the 3σ or even 4σ threshold would have a value of 0.001 pixels which would indicate that well over 99.99% of the measurement differences are within this ± 1 milli pixel range.

5.3.2 Simulation Results - Photon Noise

I now add photon limited shot noise to each of the simulated data frames and repeat the analysis. I sequentially increase the photon noise in increments and examine the mean and standard deviation of the differences in each axis for each of the simulation runs. This metric will reveal the spread around the ideal case of zero difference between the injected and measured displacements, and look at the spread of measurements in each distribution. For each case, the

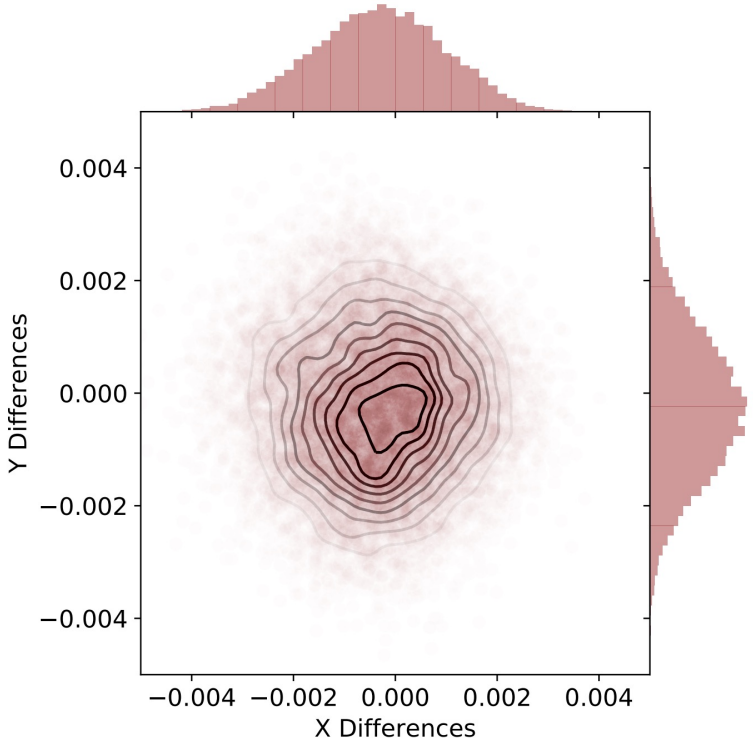
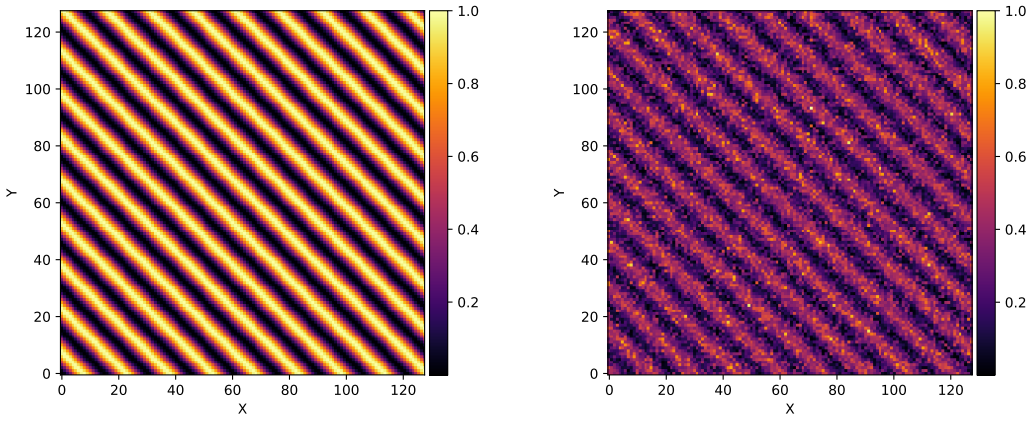


Figure 5.8: The distribution of differences between the measured and injected pixel displacements per pixel in units of pixels. A smoothed 2D Kernel Density Estimation (KDE) has been plotted to show 10 contour levels to better show the shape and off-centre position at the peak. The distribution exhibits a slight non-Gaussian distribution that is off-centre indicating a bias in the analysis procedure.

optical fringes, detector, and displacements remained unchanged, and only the photon noise was randomly regenerated between each frame as is a normal distribution whose width is some percentage of the maximum value of the fringes. Figure 5.9a shows the clean fringes with no Gaussian noise added to the frame whilst Figure 5.9b shows the fringes with the most noise added with a $1\sigma = 0.2 \times I_{max}$, i.e. noise at 20%. I attempted to increase the noise further but the 2D spatial fit started to take exponentially longer to converge on a solution, and often failed to find a good fit, so was not a practical use of time for this investigation.

To graphically demonstrate this, I have plotted the mean and spread as an ellipse for each run of the simulation with increasing photon noise, see Figure 5.10. The mean of the x and y differences is the centre coordinates of the ellipse, whilst the x and y standard deviations are the width and height respectively. These ellipses enclose the 1σ boundary of the differences distribution when approximated to a Gaussian and is a guide to how the photon noise has affected the measurement precision. Each ellipse can be thought of as a confidence contour for that set of data and allows for a quick visual comparison between the data-sets as a whole.



(a) Photon noise at 0%.

(b) Photon noise at 20%.

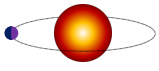
Figure 5.9: a) The optical fringes generated for investigating the pixel measurement precision with no photon noise, and b) The highest noise level used at 20%.

Table 5.2: The pixel measurement precision as a function of injected Gaussian noise, all units are in milli-pixels. The data is visualised in Figure 5.10.

Photon Noise / %	δ_x mean	σ_x	δ_y mean	σ_y
0	-1.49	1.19	-1.13	1.19
0.1	-0.33	1.35	1.37	1.35
0.2	-0.81	1.79	0.07	1.79
0.5	-0.30	4.23	0.20	4.28
1.0	0.25	8.28	-0.22	8.26
2.0	0.42	16.45	-0.15	16.41
5.0	0.17	41.60	-0.26	41.24
10.0	0.34	83.47	1.34	83.30
20.0	-1.67	167.54	5.83	167.93

Hence if an ellipse is large, it means that particular simulation poorly recovered the true pixel displacements as most of the displacement differences have a high value.

In the inset of Figure 5.10, the four lowest values of white noise are shown (0%, 0.1%, 0.2%, and 0.5%). Here it is clear that even a small level of white noise inhibits a 1 milli-pixel standard deviation, but the lowest three levels of white noise are all very similar. For the experiment, a white noise at a level of a maximum of 0.1% is therefore the target for a single set of data and will also be the nominal value of white noise when used in the simulation from here onwards. The data is tabulated in Table 5.2 where the effect of increasing photon noise rapidly decreases the overall precision of the measurements. Interestingly, the mean of the distributions rarely deviates from $(0, 0)$ by more than a milli-pixel which indicates the least-squares model is fairly robust even at moderate levels of photon noise.



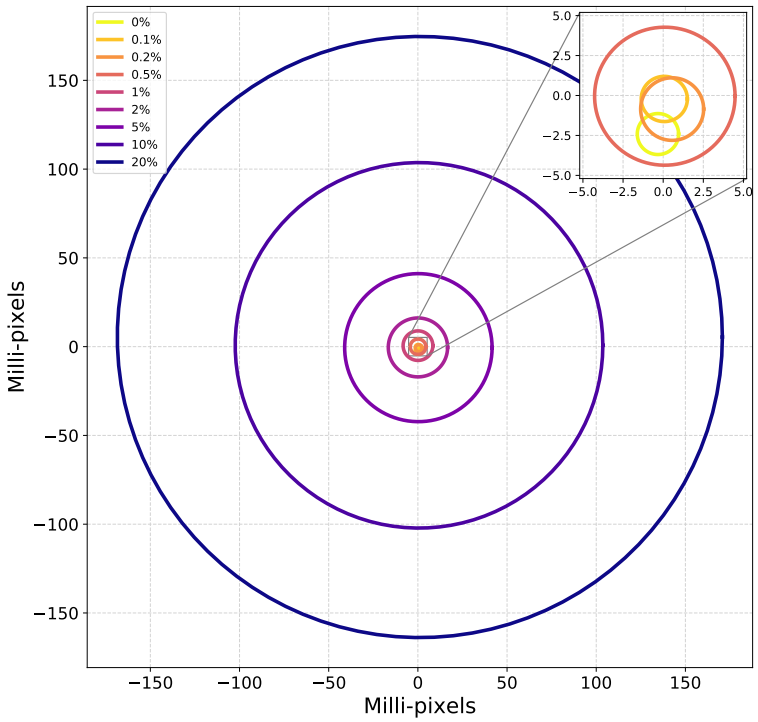


Figure 5.10: The ellipses that contain the 1σ boundary of the pixel displacement measurements for successive runs of the simulation with increasing white noise indicated by the % value. For each run, the differences between the measured and injected displacements were calculated, and the ellipse is drawn to enclose 1σ of those measurements.

5.3.3 Increased Number of Measurements

I now investigate how a different number of measurements (2D frames) during the phase sweep affects the pixel measurement precision. This will guide the experiment and give an indication of the duration of stability that is required. For example if the simulation only requires a few frames to fully describe the 1D temporal response then the data can captured in short runs. However, if the simulation dictates that many hundreds of frames are needed, then the slow temperature drifts of the environment may cause drifts between the fiber baselines and the CCD. Should these drifts be larger than 1 milli-pixel and in the plane of the CCD, the analysis will register this *as* pixel displacements.

For this test, I again set the CCD size to be 128×128 pixels, the white noise is set to zero initially, the pixel displacements are drawn from a distribution of $1\sigma = 0.05$ pixels, and the optical parameters of the fringes are set to create the fringes as seen in Figure 5.9a. I run the simulation for the following number of phase steps: $10, 20, 50, 100, 200, 500, 1000$, and in each case I note the distribution of differences between the measured and injected pixel displacements.

Table 5.3: The pixel measurement precision as a function of the number of phase-steps used for the 2D spatial fit where no photon noise is present. All of the measurements are in units of milli-pixels. $\bar{\delta}_{x/y}$ is the average of all of the differences between the measured and injected displacements, $\sigma_{x/y}$ is the standard deviation of the differences distributions.

$N_{measurements}$	$\bar{\delta}_x$	σ_x	$\bar{\delta}_y$	σ_y
10	1.37	2.22	0.20	2.22
20	0.59	0.58	0.34	0.86
50	-1.04	1.10	-0.46	1.10
100	-0.77	0.77	0.38	0.77
200	0.16	0.69	0.06	0.69
500	-0.20	1.19	0.47	1.19
1000	0.40	1.02	0.02	1.02

As in Table 5.2 I have listed the results of the mean and 1σ errors of the differences distributions but this time as a function of number of measurements during the phase sweep, see Table 5.3. For this data, the photon noise was turned off to investigate the effect of the increasing number of measurements in a noise-free environment. From this data it is clear that the 1D temporal fit is actually fairly resilient against lower numbers of measurements so long as they evenly cover the phase space which is explicitly set in the simulation. It appears that increasing the number of measurements from 100 to 1000 has not had a significant effect on the precision. This is desirable from an experimental perspective as this significantly shortens the time needed to collect the data to obtain a similar precision on the pixel displacements. A value of $N_{measurements} = 200$ will be taken as a maximum amount of frames required to adequately sample the fringes in phase-space.

I repeat this analysis but with a 1% level of photon noise randomly added to each frame. Table 5.4 displays these results. Here the effect of the Gaussian noise is evident for the series with fewer phase steps. However, as before, increasing the number of measurements past a few hundred does not drastically increase the precision of the displacements. This is likely an artefact of the Least-squares 1D temporal fit which is expected to decrease the standard deviation as $\sqrt{N}$.

5.3.4 Changing the Fringe Periodicity

In this test I keep the CCD and optical parameters the same as the previous tests, but change the fiber separation to decrease the periodicity of the fringes from many pixels to just a few. It is anticipated that smaller fringes result in more precise measurements as a given pixel displacement results in a larger measured phase difference in the 1D temporal fit which is easier to detect. However, if the fringes are too small they will be under-sampled in the 2D spatial

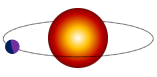


Table 5.4: As per Table 5.3 but with 1% photon added to each frame. All of the measurements are in units of milli-pixels.

$N_{measurements}$	$\bar{\delta}_x$	σ_x	$\bar{\delta}_y$	σ_y
10	0.90	3.02	-0.41	3.02
20	0.12	1.91	-0.28	1.91
50	0.12	1.73	-0.82	1.74
100	-0.26	1.86	1.71	1.85
200	-0.26	0.78	0.31	0.77
500	-0.48	1.07	0.09	1.07
1000	0.34	0.95	0.12	0.95

domain which will inhibit the spatial fit. Finding this balance will determine the geometry of the optics for the experiment. For these runs I initially turn off the photon noise and conduct 100 phase-steps of measurements, see Table 5.5. With these results it is clear that the analysis handles different fringe periods well with narrow distributions on all of the data sets. There is an anomaly at $s = 2$ mm where the distributions are abnormally large and do not fit the rest of the results. This was investigated but no source of this error was found, it could be a particular bug with the fitting routine that only happens at this particular combination of optical geometry as this anomaly was repeatable whilst the other results were repeated and also did not vary off from the reported results.

When 0.1% photon noise is added, see Table 5.6 the fiber spacing now has a much large impact on the distributions. The smaller separations exhibited a much larger error on the differences between the measured and injected displacements, but this improved with decrease fringe period. Here, the optimum was found at 5 mm onwards. The fringe period appears to have varied between the two data sets for a given fiber spacing. This is because the fringe periodicity is a measured parameters from the image by performing a 2D FFT which is affected by the presence of white noise. The fringe period is never actually used for any useful calculation of the pixel displacements, they are noted here for comparison only.

5.3.5 Varying the Fiber Angle

In this section I incrementally rotate the fibers relative to the CCD to investigate the measurement precision at different angles. It is anticipate that on average $\theta = 45°$ as this configuration will probe both the x and y displacements with equal amounts. However, there will be many pixel displacements that are nearer to perpendicular with respect to the CCD axis, and other fringe angles may measure these better. In practicality, multiple angles will be used and pixel measurements will be summed in quadrature from many sets to obtain a better estimate of each pixel position measurement. Angles of 0° and 90° are excluded as the 1D temporal fit will

Table 5.5: The pixel measurement precision as a function of the fiber spacing and thus fringe periodicity. For these results the photon noise was set to 0. All of the measurements are in units of milli-pixels.

Fiber Spacing / mm	Fringe Period / pixels	$\bar{\delta}_x$	σ_x	$\bar{\delta}_y$	σ_y
0.5	106.40	0.94	2.45	-1.46	2.04
1.0	52.80	1.03	1.16	0.55	1.09
1.5	34.96	-0.18	1.20	-0.76	1.19
2.0	28.82	-0.63	4.29	1.34	4.30
2.5	21.53	1.09	1.06	-0.21	1.08
5.0	10.69	-0.44	1.08	-1.03	1.08
7.5	7.09	-1.63	1.26	-0.12	1.23
10.0	5.32	-0.93	1.25	-0.71	1.25
20.0	2.66	0.97	1.44	-0.33	1.41

Table 5.6: As per Table 5.5 but with 1% photon added to each frame. All of the measurements are in units of milli-pixels.

Fiber Spacing / mm	Fringe Period / pixels	$\bar{\delta}_x$	σ_x	$\bar{\delta}_y$	σ_y
0.5	106.40	-1.92	27.60	27.04	24.24
1.0	52.80	-1.82	4.87	2.45	4.84
1.5	34.96	-0.39	2.87	-0.03	2.89
2.0	28.82	-0.23	4.96	1.30	4.98
2.5	21.53	-0.56	2.11	-0.89	2.07
5.0	10.69	-0.96	1.24	0.07	1.24
7.5	7.15	0.80	1.43	-0.93	1.42
10.0	5.32	-1.75	1.23	-0.28	1.22
20.0	2.66	-0.90	1.38	1.20	1.38

fail if there is not at least one complete *spatial* period in a given axis on the detector as the wavevector can not be computed without this.

I again use the optical set up as before and only vary the angle with zero photon noise on data. The results are summarised in Table 5.7. Interestingly the precision has generally increased for either shallow or steep angles which could be due to the algorithm being more sensitive to displacements that lie nearer the wavevector, see Figure 5.4. However, for a set of normally distributed pixel displacements the expectation is that all fringe angles have an equal proportion of displacements near that specific wavevector than any other. Another possible explanation is that the nearer to 45° the wavevector, the spatial fit is less sensitive to a given change in angle which leads to reduced pixel positional precision.

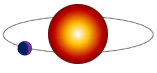


Table 5.7: The pixel measurement precision as a function of fiber angles relative to the CCD. All measurement units are in milli-pixels.

Fiber Angle / Degrees	$\bar{\delta}_x$	σ_x	$\bar{\delta}_y$	σ_y
10	-0.04	0.38	-0.4	0.19
15	-0.02	0.32	-0.15	0.39
30	-0.17	0.58	0.19	0.50
45	0.19	0.61	0.13	0.61
60	-0.41	0.57	-0.28	0.53
75	-0.24	0.35	0.02	0.38
80	0.16	0.55	0.06	0.18

Table 5.8: The pixel measurement precision as a function of simulated CCD size. The CCD is a square of size indicated by the first column. All measurement units are in milli-pixels.

CCD Size / Pixels	$\bar{\delta}_x$	σ_x	$\bar{\delta}_y$	σ_y
128	1.08	1.10	-0.57	1.10
256	-0.30	0.72	-0.71	0.73
512	-0.79	4.45	-1.00	4.45
1024	-0.09	0.98	0.31	0.99
2048	-0.30	5.14	0.01	5.14

5.3.6 Increasing the CCD size

This test is to vary the size of the simulated CCD used. It might be beneficial to analyse the real data in blocks, such as the stitching boundaries, as the 2D spatial fit computation time scales with the *area* of the CCD used, plus it can be easily parallelised. However, the 2D spatial fit assumes a Gaussian distribution of pixel displacements which becomes more accurate as more pixels are included in the spatial fit.

For this test, I use the optical parameters as recommended from Sections 5.3.2 and 5.3.4, i.e. a white noise of 0.1% and a fiber spacing of s =5 mm. The propagation distance z has remained a constant 1.5 m throughout the simulation and is unchanged here and $\theta = 45$. Also, as I only saw marginal improvement with high $\mathrm{N}_{\mathrm{measurements}}$, I keep this at $N = 100$.

I create CCDs of side lengths 128, 256, 512, 1024, and 2048 pixels. In each case I again look at the distribution of differences between the measured and injected displacements and tabulate the results in Table 5.8. Interestingly, CCD sizes of 512 and 2048 produced much broader distributions than the other sizes. This could be a one-off anomaly with this particular configuration of optics and displacements as the test was repeated which returned very similar results. This could be early justification that the CCD should be analysed in smaller blocks.

Table 5.9: The pixel measurement precision as a function of the 1σ width of the injected pixel displacements when no photon noise is present. The measurements, columns 2-5, are in units of milli pixels (mpx). The last column is the fraction of the average of the differences distribution width ($\sigma_{\text{av}} = \frac{\sigma_x+\sigma_y}{2}$) and the injected displacement width, σ_p. A smaller number indicates a more precise measurement of that distribution.

$1\sigma_p$ / mpx	$\bar{\delta}_x$ / mpx	σ_x / mpx	$\bar{\delta}_y$ / mpx	σ_y / mpx	$\sigma_{\text{av}}/\sigma_p$
0.1	0.000026	0.019	0.0000073	0.019	0.19
0.2	0.0000091	0.020	0.000070	0.020	0.1
0.5	0.000084	0.02	0.000086	0.021	0.04
1.0	0.000095	0.022	0.00047	0.023	0.02
2.0	0.0013	0.023	0.00050	0.022	0.01
5.0	0.0020	0.028	-0.013	0.027	0.0056
10.0	-0.00022	0.042	-0.021	0.050	0.0046
20.0	0.04	0.12	-0.038	0.14	0.0065
50.0	0.02	0.19	-0.002	0.20	0.0038
100.0	-0.27	0.40	0.20	0.42	0.0041
200.0	-0.69	5.74	0.147	9.874	0.029
500.0	-1.64	406.07	2.77	408.90	0.81

5.3.7 Scales of Pixel Displacements

The analysis pipeline has now shown the ability to detect pixel displacements in a variety of situations but so far only drawn from a Gaussian distribution of $\sigma = 0.05$ pixels. I now change the distribution of pixel displacements to see how the accuracy and precision changes. I generate a set of pixel displacements from various distributions ranging from $\sigma = 0.0001 - \sigma = 0.5$ pixels. Whilst it is not expected that a real CCD has displacements greater than even 0.01 pixels, it serves as a test of the pipeline in extreme scenarios. In this case I use the optimum results from the previous sections to best test the measurement precision at the lowest size of distributions and so opted for a CCD size of 256×256 pixels, fiber spacing of 5 mm, and fiber angle of 60° and I have turned off the photon noise.

When no photon noise is present, the measurement precision of the analysis reaches the level of 0.019 milli pixels. This result demonstrates the potential robust precision of the technique given some idealised circumstances such as in these simulations, but is not a physically sensible answer as this equates to a precision in absolute terms of 285 pm on the CCD, or the width of a single silicon atom on the surface of the detector.

The test was repeated with 0.1% photon noise added, as shown in Table 5.10. Here, there is a clear floor reached in precision regardless of the magnitude of the injected displacements as the width of the differences distribution does not reach lower than 0.82 milli pixels. To further increase the precision, multiple sets of measurements would need to be combined in quadrature.

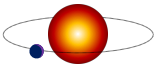


Table 5.10: The pixel measurement precision as a function of the 1σ width of the injected pixel displacements when 0.1% photon noise is present. The measurements, columns 2-5, are in units of milli pixels (mpx). The last column is the fraction of the average of the differences distribution width ($\sigma_{\text{av}} = \frac{\sigma_x+\sigma_y}{2}$) and the injected displacement width, a smaller number indicates a more precise measurement of that distribution.

$1\sigma_p$ / mpx	$\bar{\delta}_x$ / mpx	σ_x / mpx	$\bar{\delta}_y$ / mpx	σ_y / mpx	$\sigma_{\text{av}}/\sigma_p$
0.1	-0.0016	0.82	-0.0056	0.82	8.2
0.0002	-0.010	0.82	-0.0050	0.82	4.1
0.5	-0.0049	0.82	0.0028	0.82	1.64
1.0	0.0017	0.82	-0.0013	0.82	0.82
2.0	-0.0090	0.82	-0.013	0.82	0.41
5.0	0.0068	0.82	-0.020	0.82	0.164
10.0	-0.017	0.82	-0.0020	0.82	0.082
20.0	-0.031	0.85	-0.021	0.83	0.042
50.0	-0.090	0.87	0.48	0.85	0.017
100.0	-0.16	1.11	-0.65	1.22	0.011
200.0	-0.41	1.33	-0.49	1.53	0.0072
500.0	2.16	398.68	-1.47	408.34	0.81

5.3.8 Simulating Stitching Boundaries

The last part of the simulation is to generate the effect of stitching boundaries. These are an artefact of the photolithography process where CCD is is made from an array of smaller CCD sections which is stepped across the material. During this process there may be some misalignment between the blocks which causes a row or column of pixels to have a slightly larger or smaller area than expected. This is usually visible in a flat-field image as those pixels have collected more light for a give exposure, see Figure 4.2.

To simulate this, I create a block of a given size that is smaller than the CCD, e.g. a 512×1024 pixel block for a 2048×2048 CCD (which is the case for the specific CCD used in the experiment). For each pixel in the block, I displace every pixel inside it by a fixed amount in x and y. This process is then repeated across the CCD until each pixel in a given block is given a displacement of that block. The block displacements are drawn from a normal distribution set of $1\sigma_{\text{block}} = 3\sigma_{\text{displacements}}$, i.e. the block stitching effect is 3 times larger than the individual pixel displacements on average. An example of the effect of a block displacement is shown in Figure 5.11.

I assume that the block *only* contains a linear displacement, whereas in reality it could also be slightly rotated or warped. The purpose of this part of the investigation is to see how the analysis handles a block displacement but not to correct for it. If the analysis fails, then I know I need to analyse each block independently or adjust the pipeline to handle block stitching.

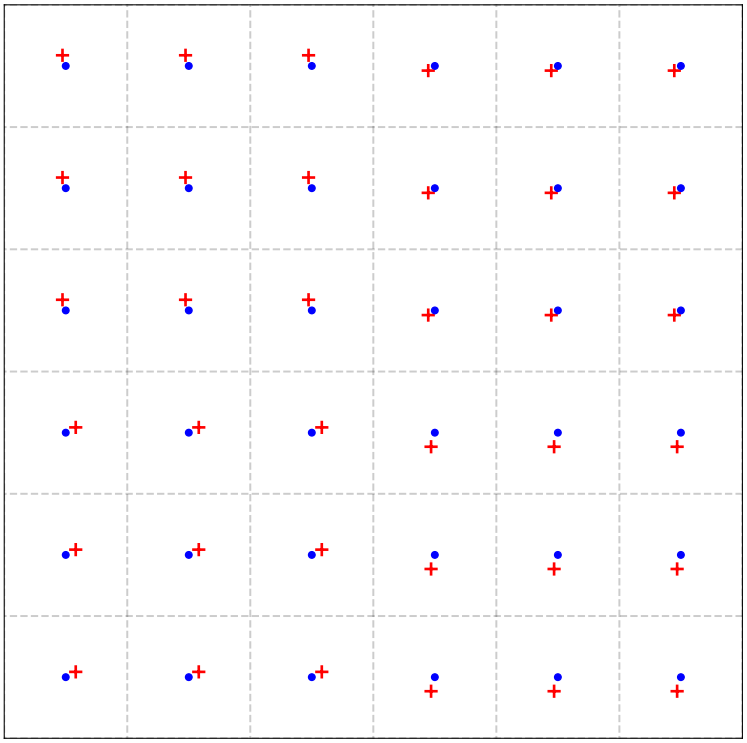


Figure 5.11: An example of how a block displacement leads to stitching boundaries in the simulation. In this case, a 6×6 pixel CCD is comprised of 4 blocks of 3×3 pixels. Each pixel in the block is displaced as a whole, and the block displacement is drawn from a normal distribution.

To see the effect on a global displacement map, I have plotted the X and Y displacements of a 256-pixel detector with 2×2 blocks of equal size. The block displacements are drawn from a distribution of $1\sigma = 0.03$ pixels, and each pixel within a block is then given an additional random displacement drawn from $1\sigma = 0.01$ pixels. Figure 5.12 shows how the stitching is clearly visible over the random per-pixel displacements. For this particular configuration, the distributions of injected displacements are no longer simple 2D Gaussians, instead it will be a set of overlapping Gaussians (multi-modal 2D distribution) of the same size as the number of individual blocks created. For the displacements shown in Figure 5.12, the injected block displacement values were $x = (-0.0116, -0.0335, -0.0218 - 0.0031)$, $y = (0.0141, 0.0081, -0.0055, -0.0094)$. These coordinates are marked on Figure 5.13.

To assess how well the analysis has handled the presence of block stitching, I reproduce Figures 5.6a and 5.6b but now include the block displacements as seen above. In Figure 5.14a it is possible to discern two well correlated distributions indicating the analysis has handled some of these particular block displacements well, whereas in Figure 5.14b the distributions are more smeared an undistinguished. The actual values for the Y block displacements happen

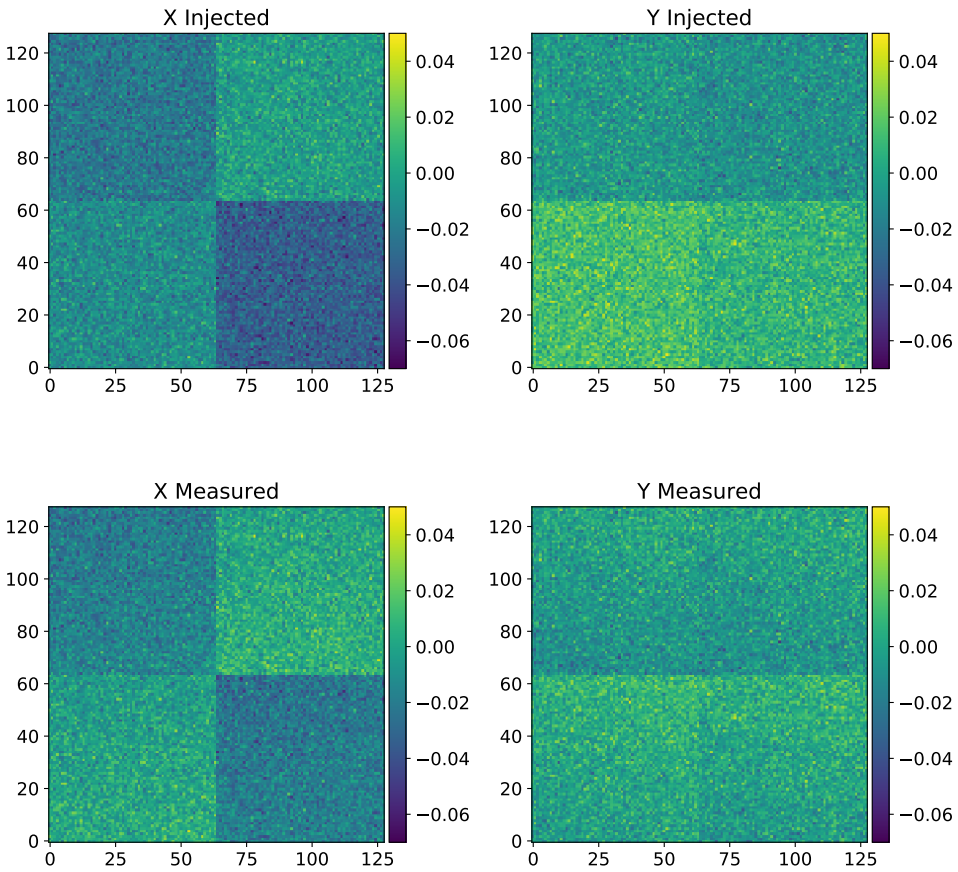


Figure 5.12: The X and Y displacement maps of a 128×128 pixel CCD with 2×2 equal block sizes. The colour-map is the value of each pixels displacement along a given axis and is equal across all the plots. The top row is the inject pixel displacements in X and Y, and the bottom row is the measured pixel displacements in X and Y.

to be in two close pairs rather than spread out like the X values which could have affected the results here.

5.4 Simulation Conclusions

The optical fringes simulation and analysis pipeline have proven to be useful in determining the general configuration of an optical experiment and in giving a benchmark of expected performance. I have verified that the experiment can be constructed with reasonable confidence in that the results do not suggest that the geometry needs to be of any extreme size or dimension. The analysis simulation has also verified that the shape and size of the data, i.e. how many

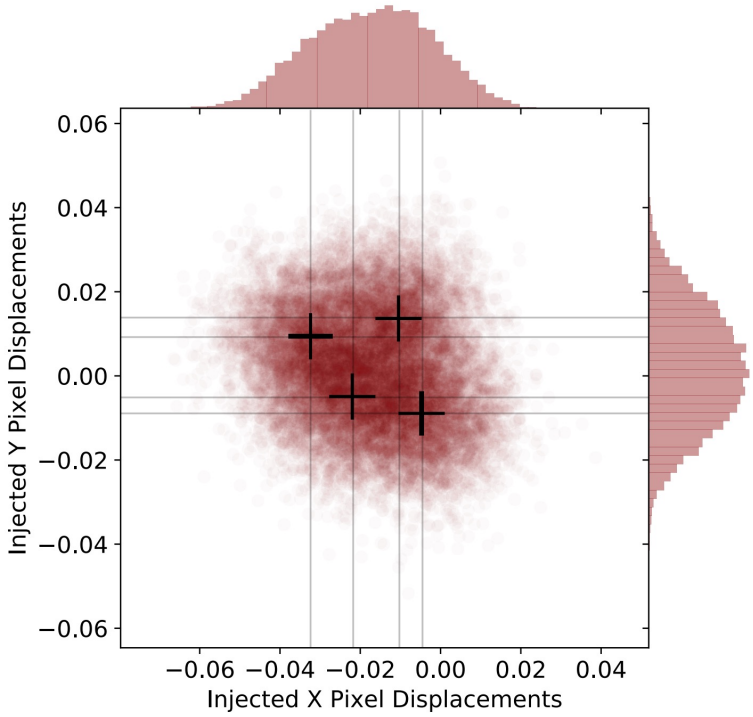
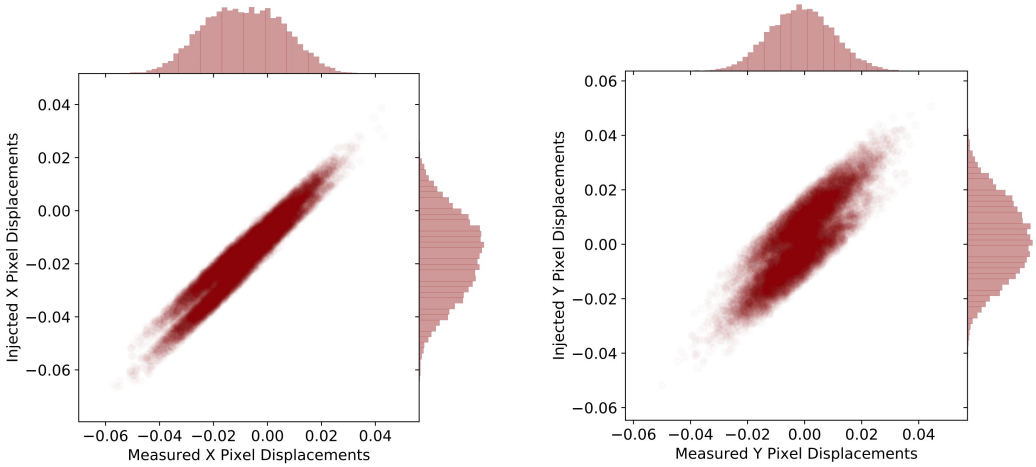


Figure 5.13: The distribution of injected pixel displacements for a simulated CCD when block displacements are also present. The black crosses mark the values of the each block displacement which will also be the centre of a 2D Gaussian distribution of pixel displacements for that block.



(a) Injected and measured displacements in x

(b) Injected and measured displacements in y

Figure 5.14: a) the bivariate distribution for the measured and injected horizontal (x) pixel displacements from a simulated CCD. b) the same for the vertical (y) displacements where both distributions now contain the block stitching displacements as well as the normally distributed displacements per pixel.

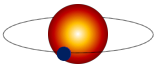


Table 5.11: The summary of the optimum parameters from the CCD analysis simulation. These results will be used to dictate the construction of the optical experiment.

Parameter	Optimum Value	Comment
$N_{\text{measurements}}$	100-200	Might be time-limited
Photon Noise target	1%	Will require CCD calibration
Fiber Angle	60°	Multiple will be used
Fiber Spacing	5-10mm	Multiple will be used

frames I need to capture in a given run, is also achievable in a realistic time scale.

The photon noise test showed that the analysis is able to handle even a large amount of white noise on each frame and still find accurate results on the pixel displacements. I predict that the optical experiment will need to have noise of below 1% which is where the analysis produced the most accurate displacement measurements. Analysing a larger CCD did not drastically improve the measurement precision which is promising if the analysis needs to be conducted in smaller blocks for any reason. This is likely due to the low noise level used in the simulation, and I anticipate that including more frames would be beneficial in a regime of high noise.

The fiber spacing and angle variation showed that an optimum lies at around 5 mm and that any angle produced good results if the fringe periodicity is low, particularly shallow or steep fringe angles. This aids in designing the fiber baseline mounting block. The results are summarised in Table 5.11. Whilst this simulation factors in many different aspects of a real experiment, some shortcuts had to be made in order to end with a complete pipeline in a sensible time-frame. In no particular order, here are some of the extensions and next steps that could be taken to further improve the analysis technique.

The simulation could be used to to fully optimise the experiment to *tell* the user how exactly to construct the optics and sample the fringes to result in the best possible measurement precision of a particular detector. The parameters λ, θ, s, z, w, and $\mathrm{N}_{\text{measurements}}$ can be optimised by feeding the result of the precision back into the simulation and looping again to further optimise the precision. This would require a near full re-write of the pipeline architecture, but could provide useful insight for future experiments. The model assumes that all of the pixels are either displaced from a normal distribution, or in a block that contains a normal distribution. Whereas it is possible that the displacements form large scale patterns such as those found in Shaklan et al. (1995), that result from fabrication processes. It would be fairly trivial to inject these kind of displacements in a simulation, but fairly difficult to attempt to model any arbitrary displacement pattern. The simulation also assumes that each pixel is the same size and only displaced, it is possible that the pixel size also varies on some small scale. This could result in an appearance of a non-uniform flat-field even in the idealised case of zero photon noise.

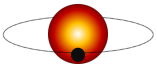
Chapter 6

The Experiment to Measure the Pixel Positions of the HARPS3 CCD

'Computers aren't the thing. They're the thing that gets us to the thing.'

Joe MacMillan

In this chapter I outline the steps taken to construct an experiment capable of generating optical fringes, as described in Chapter 5, and present the analysis of the measured pixel displacements of two optical detectors. The design and construction of this experiment was guided by the results of the simulation in Chapter 5 and augmenting the strategies used by Shaklan et al. (1995) and Crouzier et al. (2012) to better fit the needs of the HARPS3 experiment. I first detail the choice of components and the justification for their specifications as determined by Table 5.11, and discuss the range of optical fringes possible with this set up. I also designed and manufactured some custom machined items for the fiber control and CCD mount with fringe control in mind. Secondly I discuss the various component replacements that were necessary after some of them did not meet the requirements, and how that affected the data collection. Thirdly I discuss the data collection strategy and how I mitigated environmental effects, and lastly I discuss the results from the experiment.



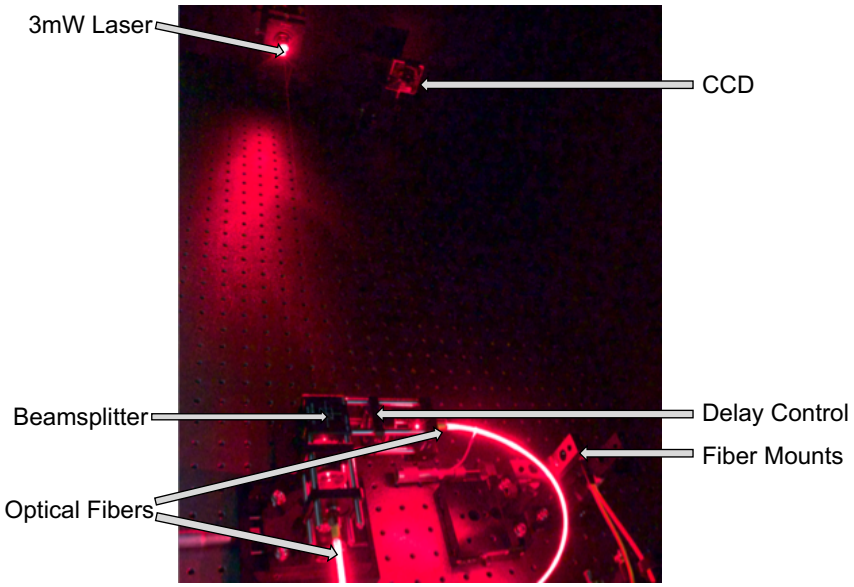


Figure 6.1: The components of the first experiment to generate optical fringes. The distance from the fiber mount to the CCD is approximately 1 m.

6.1 A Simple Interferometer

The first iteration of the optical experiment served as a proof-of-concept for fringes illuminating a detector, and as a method to analyse the effects of the environment.

It consisted of a 3 mW open-air diode laser at $\lambda = 635$ nm aimed at an optical 50:50 beamsplitter. Each channel of the beamsplitter contained a lens that collimated the beam into a fiber. The two fibers were then placed facing a distant detector. One of the lenses was placed in a translating stage that allowed it move along the beam axis which creates a variable delay line to control the phase of the interference pattern. This setup is pictured in Figure 6.1.

Along with this setup, I placed a small environmental monitor to measure the ambient air temperature, pressure, and humidity. To monitor the environment and see the effect of any changes on the optical fringes, I captured a long series of simultaneous environment data and fringes. For each fringe, I fitted a simple sine-wave to the middle row of the image at each step. If the parameters of the fit (amplitude, period, phase) changed by some significant amount over a data collection run and this correlated with the environmental data, this would indicate that the fringes are not stable and more mitigation is needed. This data was collected *before* the detailed simulation of the optical experiment was conducted, so there are no target tolerance levels. This was a simple experiment to monitor the environmental affect on optical fringes to act as a baseline of which parameters are mostly affected by some change, and to act as a starting point on an investigation into mitigating these variations.

The 72 hour series is shown in Figure 6.2, a short exposure of 0.1 s is captured every

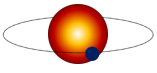
minute. The daily temperature variations within the optical blackout box are clear from the 4°C, 24 h variation. Temperature fluctuations cause the biggest variation in the phase of the optical fringes due to the expansion and contracting along the beamline acts as an optical delay line. For a 635 nm laser, a path length change of this amount will result in a complete period shift in the fringes. Even a small physical change of just, say, 25 nm will result in a phase shift of ~5% of a period on the detector. The temperature variations here have clearly affected the phase which means some part of the optical beamline is changing. From Figure 6.2, when the temperature gradient is low, the phase is relatively stable but for all other temperature gradients a large phase sweep of many periods is seen. This is undesirable for an experiment where phase control is critical.

Figure 6.2 also shows that the amplitude of the fringes were, however, relatively stable during this period which suggests that the changes in path difference of the beamlines did not affect the delicate coupling from air to fiber. Constant amplitude of fringes will mean that any spatial fit is consistent between frames as if the contrast is too low, the spatial fit can be adversely affected.

The offset (background noise) of the fringes shows some slight increase at higher temperatures. This could be extra dark current in the CCD due to the increased thermal energy, or it is possible that the blackout box was not sufficient to mitigate the day/night cycle.

With this quick and simple test, the need for environmental control and fringe stability was apparent. This particular detector could be read out at many tens of frames per second so even short term drifts did not pose a problem. However, for the full-size CCD, the read out is much slower as it scales with the area of the CCD in pixels. For a large-format 2048×2048 pixel detector, a full read-out can take as long as 30 s, which means many hundreds of phase steps to complete a cycle could take hours of data collection. The drifts on these timescales could affect the periodicity also which would render the analysis redundant as it assumes consistent fringe spacing.

It was decided at this early stage to abandon any further experimentation with this current set up, and design an experiment that offered more control and stability of the optical fringes. To aid in this design, and see the effect of the design choices on the pixel mapping results, the complete end-to-end simulation of Chapter 5 was then used. Using the optical physics of Section 4.4.1, a full optical model was created that would output images of interference fringes from a defect detector, and the analysis could be prepared on this data. Here, the effect of the design choices could be seen in real-time and the experiment can be tailored to optimise results.



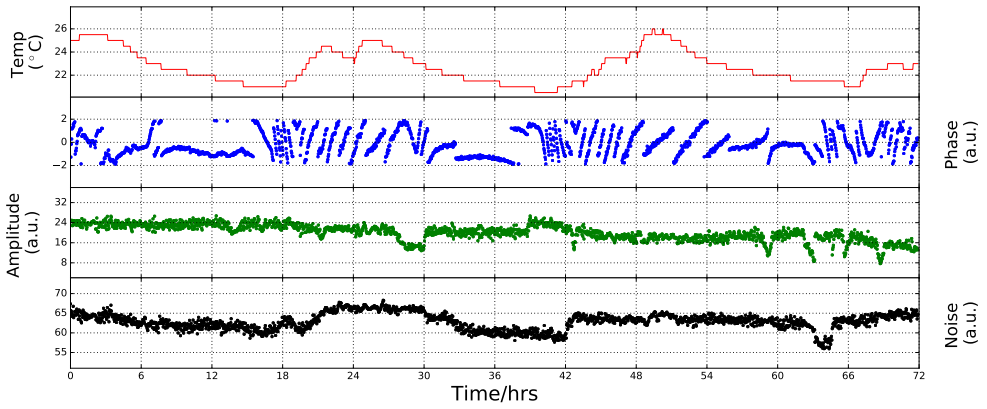


Figure 6.2: The 72 hour long data collection on the initial fringes. The fit parameters are shown as a function of time, along with the measured temperature of the black-out enclosure. The 'Phase' parameter is range is restricted hence the discontinuities. The 'Noise' parameter is the zero-offset in counts and is a measure of the background of the enclosure.

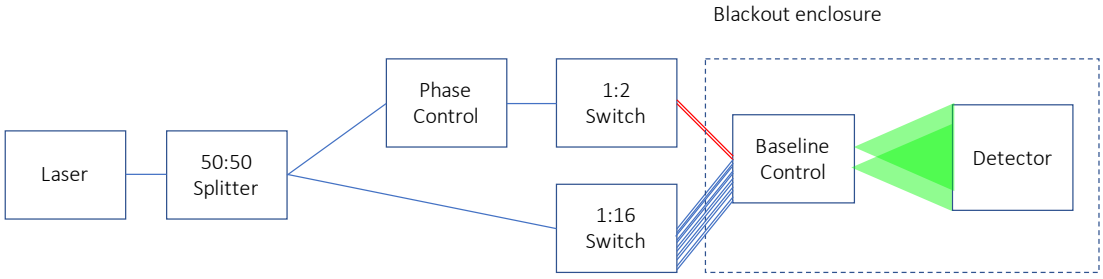


Figure 6.3: A block diagram of the experiment showing the various components discussed in Section 6.2 .

6.2 Building a Better Experiment

The parameters that define the optical fringes (w, s, z, θ, ϕ) can be controlled by different groups of components. In this section I list the design choices used to tune the parameters which yield optimum optical fringes. The overall schematic of the experiment is shown in Figure 6.3.

The fiber core diameter, w, is determined by the choice of fiber. Single mode (SM) optical fibers are necessary for interference as they only allow one spatial mode to propagate which results in minimal optical speckle and ideal interference. SM fibers have very narrow cores of just a few µm. To allow for multiple wavelengths of different lasers, I chose Nufern S-405XP which has an operating wavelength of 400 nm to 680 nm and a mode-field diameter (MFD) of 4.5 µm[a]. S-405XP is a common fiber and can be ordered 'off-the-shelf' with many connection

[a]For a Gaussian power distribution, the MFD is the diameter where the optical power is reduced to $1/e^2$ from its peak level and encompasses over 99% of the distribution.

configurations.

The fiber spacing and angle, s and θ, were optimised to be around 5 mm and 60° respectively from the simulations but only marginally, see Table 5.11. Hence an optical block was designed that can hold the fibers at multiple separations and angles from 2.5 mm to 10 mm and 0° to 90°. The block was designed to to hold 16 fibers simultaneously such that data-cubes from multiple angles and baselines can be taken without disturbing the experimental set up. An annotated diagram of the custom block is shown in Figure 6.4, this was the actual engineering drawing supplied to the workshop for construction. Here, the block has multiple grooves on two perpendicular sides where the fibers are mounted. There are also grooves on the other 2 sides of the block which were machined for a potentially narrower fiber option that ultimately was not used. The block itself has a large heavy mount for the optics bench where the block can sit at increments on 45°. The block was specified to be manufactured from aluminium. The linear coefficient of thermal expansion (CTE) of aluminium is $23.1 \times 10^{-6}\,\mathrm{K}^{-1}$ at 20°C. Hence, a 1°C change at typical lab temperatures will result in a fiber baseline change of just over 20 parts in 1 million. For a 5 mm fiber spacing, this will increase by just over 10 µm. As fringe periodicity is inversely related to fiber spacing, a change of this magnitude will decrease the fringe period by 2 milli-pixels. Hence, thermal stability of less than 1°C over a data collection period is required, or the optical components will have to be re-made from lower CTE materials.

To split the laser light into two equal intensity channels a fiber coupler (sometimes referred to as a fiber splitter) was used. These are passive fiber-optic devices which are manufactured by splicing two fibers together at a very shallow angle such that the mode couples to both. I initially made contact with many manufacturers as an optical, broadband, SM fiber splitter was not readily available. Initially I purchased such a device but found the splitting ratio as a function of wavelength varied wildly across the 400 nm to 700 nm range. As a stop-gap, a single wavelength splitter was used to match the wavelength of the laser and could be replaced at a later date should other wavelengths be needed.

The length of the optics bench used constrains the propagation distance z. I also wanted to enclose the entire optical path in a black-out box to mitigate stray light from external sources. From the simulations, $z = 1.5$ m proved to be a useful distance as it is an optimum trade off between fringe periodicity, photon counts, and making sure the detector sits at the peak of the central lobe of the sinc envelope. Hence I constructed a black-out enclosure of dimensions $1.5 \times 0.6 \times 0.6$m. The actual propagation distance is, then, slightly less than $z = 1.5$ m as both the fiber mount and CCD mount must fit inside the enclosure. One of the long walls of the black-out box was given a sliding door such that components can be easily accessed. The black-out box also has the added benefit of reducing air turbulence that could negatively affect the fringe stability as seen by the detector.

Lastly, the block that controls the baseline can be populated with multiple fibers and a switch can be used to choose which pair of fibers are illuminated. I use two switches, one for each of the initial channels, that control the baselines. The first is a 1 channel to 2 channel switch where one of the outputs are mounted in x and the other in y on the block. The other switch is 1 channel to 16 channels, where 8 are placed on each axis. This means that I can choose any pair that satisfies the choice of any one from each of the switches. This is graphically demonstrated in Figure 6.5.

6.2.1 Controlling the Phase

The phase shift is a key parameter to control as it allows for the fringes to be uniformly swept across the detector such that the 1D temporal fit has a well sampled time series. After investigation, the options for phase modulating include: an optical delay line, an electro-optic phase modulator, and a fiber stretcher. Each option has inherent pros and cons for the purpose of fringe control.

The optical delay line is a tried-and-tested 'off-the-shelf' piece of equipment. It consists of two adjacent fiber ports that each face a mirror at 45° such that the light leaves one port, bounces off the two lasers in a U-shape beam, and re-enters the second port. The mirrors are mounted on a precisely controlled stage and can move in increments down to just a few nm. As the wavelength is 523 nm, a delay shift of this quantity corresponds to a full 2π sweep of the fringes in the direction of the wavevector. Hence, such a delay line could sample a single period in around 20 steps which is likely adequate for the purposes as investigated in Section 5.3.3. However, the beam is launched through the air which makes it susceptible to turbulence. Plus, any uncharacterised temperature variations could cause the delay baseline to expand and contract which would induce unwanted delays and move the fringes. Also, launching and collecting a beam into an SM fiber is extremely delicate as the core diameter is just 4.5 µm. Any slight deflection or deviation off perfect alignment will drastically reduce the amount of coupled light.

The electro-optic phase modulator is a small piezo-electric crystal which expands and contracts as a voltage is applied. The crystal has such a narrow range of expansion which means it is ideal for delicate phase control of an interference pattern. The example unit in mind was the Thorlabs EO-PM-NR-C4 which has high transmission for 400 nm to 600 nm and the full voltage range gives just over 2π phase modulation across the entire range. The phase-modulator is also resistant to temperature changes which means the applied voltage is likely the only source of path length difference. However, as with the delay line, the light needs to be launched out of and back into a narrow core SM fiber. But, unlike the delay line, the length of open beam can be controlled by the user to be as small as possible to reduce the risk of bad coupling, and

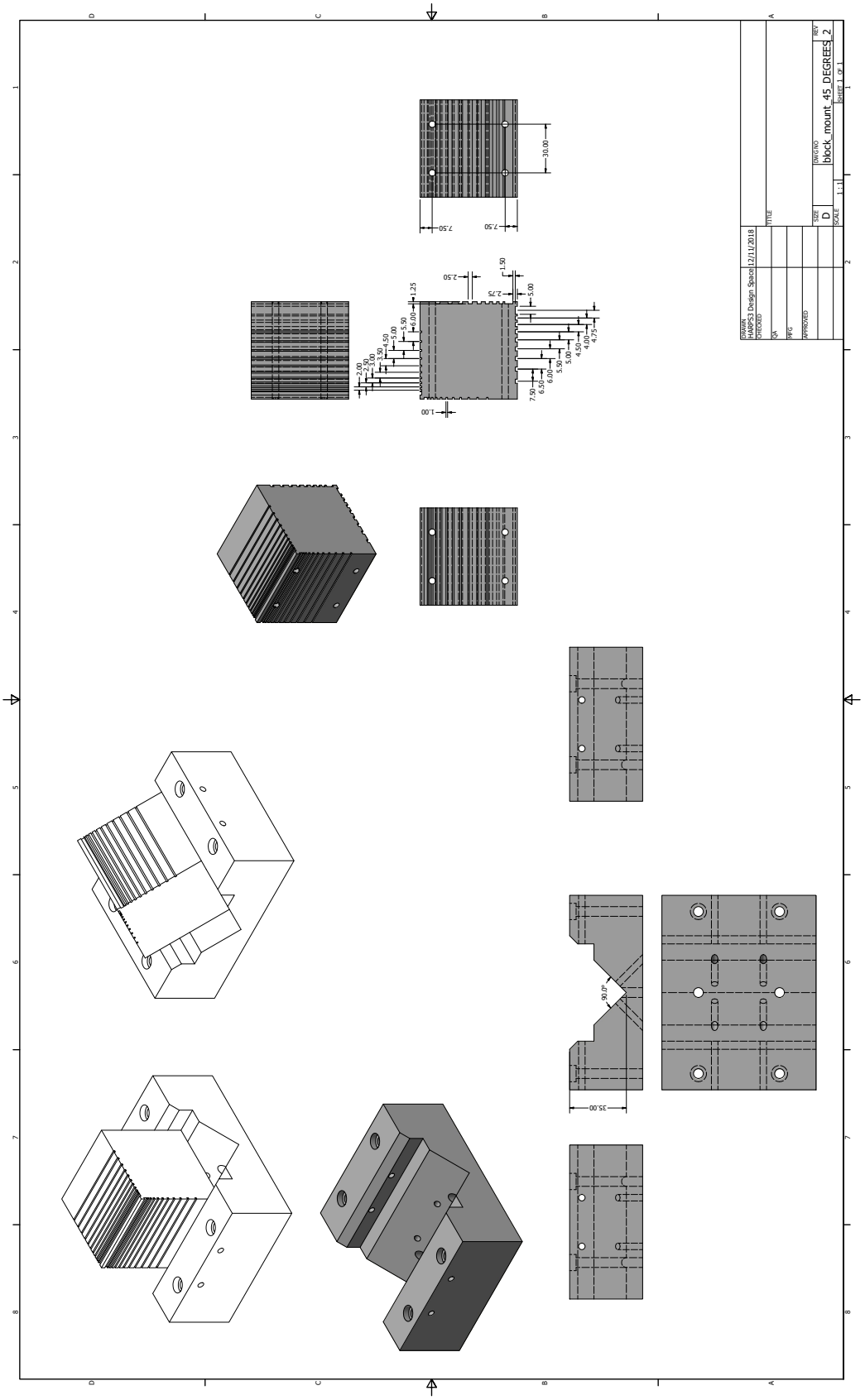
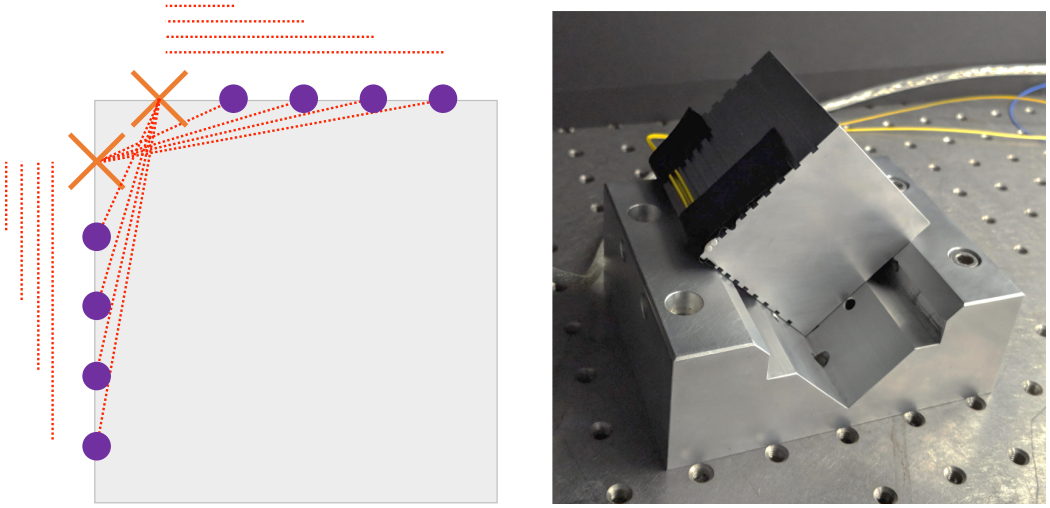


Figure 6.4: The CAD drawing for the fiber block and the block mount that is bolted to the optics bench.



(a) Fiber block baseline schematic. (b) Fiber block photograph.

Figure 6.5: a) The complete set of possible baselines as seen by the CCD for two switches where one is 1:2 and one is 1:8. Any pair of fibers can be selected provided it contains one orange '×' and one purple '∘' position. In reality the second switch is 1:16 so there are twice as many '∘' options, but I have reduced it to 8 for clarity for this schematic. The red dotted lines represent the relative wavevectors that are produced from each pair. There are 32 available pairs and the block itself can be mounted at 45° intervals. Additionally, the '×' fibers can be placed at any of the positions which greatly increases the number of available baselines. b) a photograph of the manufactured block with two fibers mounted.

the beam axis misalignments. Plus the phase-modulator can be installed in a precise optical alignment setup which are ideally suited for SM fiber coupling.

Lastly, the fiber stretcher consists of a fiber being wound around a small piezo stack, such as a cylinder, which changes in radius when a voltage is applied. The tight winds of fiber each get an additional path length as the piezo stack expands. These stacks are controllable and robust but only move by small amounts hence many winds of fiber are required for a sizeable path-length change. The fibers and the stack are subject to temperature variations so are often constructed in a feedback loop to maintain some physical displacement. However, finding a suitable fiber stretcher that has the desired specification of broadband transmission, high temperature stability, a feedback loop, and all at a reasonable price and delivery timescale was difficult.

Hence, the Thorlabs EO phase modulator was selected along with an optical alignment bench to maintain coupling. A 50:50 in-fiber splitter is used to equally split the intensity between the two channels, one of which contains the phase-modulator.

6.2.2 Estimating Optical Losses and Laser Selection

The last component to choose is the laser. The power needs to be sufficient to register nearly full well depth counts for a sensible exposure time given all of the losses of the various components in the optical path. To estimate this, all of the losses are combined from the laser to the baseline to obtain the fraction of power emitted by the fibers. Then, the propagation distance between the fibers and the detector are used to estimate the fraction of this that a single pixel collects.

The wavelength of light, λ, was chosen to be 532 nm as this is roughly in the centre of the optical range of HARPS3 (380 nm to 680 nm). However, the experiment was designed from the ground-up to allow for other wavelengths of laser to be used within the entire optical range of HARPS3. This is because the sensitive collection area of the pixel may deviate as a function of wavelength which could then mean the effective pixel position varies with λ. Photons of different wavelengths will penetrate the semiconductor material at different depths due to their energy, and if the insulating pixel boundary is not perfectly uniform throughout the material the pixel position may vary with wavelength. Hence, a pixel displacement map conducted at just a single wavelength may not provide a complete description of the detector geometry at wavelengths away from that of the laser used.

With the laser wavelength selected, the power range must now be chosen to to enable sensible exposure times and to ensure the detector is not saturated at the lowest power available. The reported worst-case losses for the splitter, phase-modulator with fiber ports, 5 meters of SP 405 XP fibers with FC/APC connectors, and two switches are 8 dB, 5 dB, 0.5 dB, and 1.4 dB respectively[b]. As decibel units are logarithmic, they can be summed for the total optical losses to achieve an attenuation of 16.8 dB. Converting to fractional power losses with

$$P = P_0 \times 10^{L/10}, \tag{6.1}$$

where P_0 is the initial power and L are the losses in decibels, results in a fractional power transmission of just 2.34%.

To estimate how much of this power a single pixel collects, I follow the procedure in Crouzier et al. (2012). The Gaussian beam emitted by a single fiber at a distance z is given by

$$I(z) = \left(\frac{\omega_0}{\omega(z)}\right)^2 I_0, \tag{6.2}$$

where ω_0 is the beam waist at the origin and $\omega(z)$ is the beam waist at a distance z, and I_0 is the initial intensity. In the far-field regime we can approximate $\omega(z) = \mathrm{NA}z$ and $I_0 = \frac{2P_{\mathrm{tip}}}{\pi\omega_0^2}$ where P_{tip} is the power of the source. This gives

$$I(z) = \frac{2P_{\mathrm{tip}}}{\pi \mathrm{NA}^2 z^2}. \tag{6.3}$$

[b] Values obtained from the various specifcation data sheets per component

Table 6.1: The major components used to construct the optical path as shown in Figure 6.3.

Component	Model	Specification	Quantity
Laser	Vortran Stradus 532	532 nm, 40 mW	1
Switch	Lightech Optical Switch	1:2 and 1:16, S405-XP, FC/APC	1 each
Switch PSU	TTi EL302RD	30 V, 2 A, 2 Channel	1
Splitter	Thorlabs TN532R5A1	50:50, FC/APC	1
Phase Modulator	Thorlabs EO-PM-NR-C4	400 nm to 600 nm	1
Phase Modulator PSU	Thorlabs HVA200	0-200V output	1
Fibers	Thorlabs S405-XP	FC/APC	18

The power received by one pixel at the centre of the beam is approximated by:

$$P_{\text{pix}} \approx \frac{2P_{\text{tip}}\text{px}^2}{\pi \text{NA}^2 z^2}, \tag{6.4}$$

where px is the side length of the square pixel. Assuming the power is uniform across the area of the pixel, which by design of the experiment it almost exactly is, the photo electron flux received by the pixel is

$$F^{e-}_{\text{pix}} \approx \frac{2P_{\text{tip}}\text{px}^2 \lambda \text{QE}}{\pi \text{NA}^2 z^2 hc} \tag{6.5}$$

where QE is the quantum efficiency of the detector at wavelength λ, and hc is Planck's Constant multiplied by the speed of light in a vacuum.

From Equation 6.4 the maximum fraction of the light collected by a pixel is around 5×10^{-15}. Converting this into an electron flux with Equation 6.5 yields approximately 1×10^6 photoelectrons per second with a laser of 1 W. For a full well depth of approximately 150,000 electrons, the exposure would saturate the CCD in just 0.15 seconds. Hence, I would need to either attenuate the laser with a neutral density filter or choose a lower power laser. For these reasons I opted for a 40 mW diode laser operating at 532 nm. I chose a Vortran Stradus 532 model with a SM FC/APC connector[c]. This laser can be operated at mW increments between 1 mW to 40 mW which allows for a broad range of exposure times without saturation, but is bright enough to allow for the case where the losses have been underestimated as it is predicted to saturate the detector at very short exposures. For a 1 second exposure at 40 mW with this, I anticipate 40,000 counts on the HARPS3 detector, roughly 1/3 saturation.

The final schematic of the experiment is shown in Figure 6.3, and the list of components used is in Table 6.1.

[c]The specification of the laser can be found at:
http://www.vortranlaser.com/images/stories/printables/Stradus%20532-40%20Datasheet%2010867.pdf.

6.3 Choice of Optical Detectors

During the experiment, two different detectors were used. The first was a Thorlabs CMOS USB camera with a resolution of 1280×1024 pixels[d]. Whilst a CMOS sensor pixel architecture is completely different from a CCD sensor and the pixel mapping results could be meaningless, this camera was useful for alignment, testing, data analysis preparation, and initial fiber imaging. A CMOS sensor has an individual amplifying circuit per pixel which massively increases the readout speed but comes at a cost of photometric sensitivity and uniformity. This easy to use detector proved essential in building the experiment.

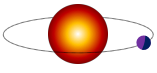
The experiment and simulation was designed for the 4096×4096 pixel science CCD of HARPS3. This CCD is to be mounted in a vacuum with a window to the spectrograph optics. The vacuum is part of a larger continuous flow cryostat to keep the detector at liquid nitrogen temperatures. During construction of the experiment it became apparent that delays to the overall HARPS3 project would mean that this detector would not be ready in time. Hence a second CCD was ordered from e2v that acted as a test detector for this experiment and also as a practice run for setting up the ESO controller that shares a similar configuration and electronic architecture. It is referred to as the 'HARPS3 Test CCD' throughout this thesis.

The HARPS3 Test CCD used for the final iteration of the experiment is the e2v 'CCD230-42 Back Illuminated Scientific CCD Sensor 2048×2048 Pixels, Four Outputs and Inverted Mode Operation'[e]. This description means the CCD is operated in inverted mode by holding the sensor at a large negative voltage and the photons result in the generation of holes in the substrate as opposed to the usual electrons. In this mode, the CCD is much less susceptible to dark current at room temperature but this comes at the cost of higher read-noise for each frame. This choice of operation mode was essential as the standard 'non-inverted' mode *requires* the use of a cryostat which was unavailable at the time. Hence, it was necessary to use a detector that can be operated at room temperature but one that shares the architecture of the HARPS3 CCD. Additionally, the PCB used shares a lot of the common electronic architecture (controller, pre-amplifier, cables) that will be used with the HARPS3 CCD. The CCD has four outputs which allows the controller to operate the CCD in four separate circuits which results in a read-out speed increase of 4×. For fringe stability on longer times scales this could be significant.

For this CCD a custom PCB was designed by HARPS3 consortium member John Ely, and manufactured externally. The PCB was specified for this particular experiment but some features would be carried over to the larger 'Science-grade' CCD sensor for the actual HARPS3

[d] CMOS data-sheet can be found at: https://www.thorlabs.com/thorproduct.cfm?partnumber=DCC1545M, and sensor data-sheet here: https://www.1stvision.com/cameras/sensor_specs/MT9M001-D.pdf.

[e] e2v CCD data-sheet at: https://www.e2v.com/resources/account/download-datasheet/3828.



instrument. The specified requirements were:

- No physical feature protrudes past the CCD front face
- All cables and mounts must be rear-mounted
- The CCD controller and cables will be common to this PCB and the full size science-grade board used for HARPS3
- The CCD will sit close to one edge of the board such that the height of the CCD can be exactly controlled.
- The board will be fastened to large optical bench mounting brackets, and as such it will need to accommodate the mounting points for this.
- The CCD socket will be low-force insertion so as to not risk damaging the CCD during installation

With these factors in mind, I designed and manufactured an aluminium PCB mount that accommodates the specifications of drawings for the PCB and those of the chosen mounts. The schematic is shown in Figure 6.6. The black right-angled brackets are stock items from Thorlabs and are used to take the weight of the board and the strain from the cables by directly bolting into the optical bench. The various mounting points up the long vertical side allows for the CCD to be mounted at heights in increments of between 10 mm to 25 mm, with the CCD being mounted at the middle of the bottom edge. The side and top panels are used to support the weight of the cables such that they do not warp or bend the board. In the rear-view, the mounting points for the various cables can be seen, these will ultimately route out the back of the black-out enclosure. The hole behind the CCD is a gap for a potential passive cooling block should the dark current of the detector be too high. The PCB has 7 mounting points to the block where spring-loaded washers are used to equalise the pressure across the surface.

The PCB mount is made from machined aluminium and as such is heavy relative to the CCD. This large mass is advantageous for mitigating potential vibrations and acting as a thermal sink. The mount will have a precise temperature probe attached to track any potential sources of temperature-dependent errors that may arise.

6.3.1 Inspection of the HARPS3 CCDs

In this small aside, I include a few photographs taken during the inspection of the HARPS3 grade 0 CCD, the actual CCD that will be installed in the instrument and used to capture the spectra of distant stars. These were taken in a cleanroom used for photolithographic processes and hence only a narrow-band of yellow light is used to illuminate the room. The detector is enclosed in a small metal box that directly protects the fragile surface. This box is mounted in a large cylindrical chamber with a vacuum seal to protect the CCD from any moisture that may

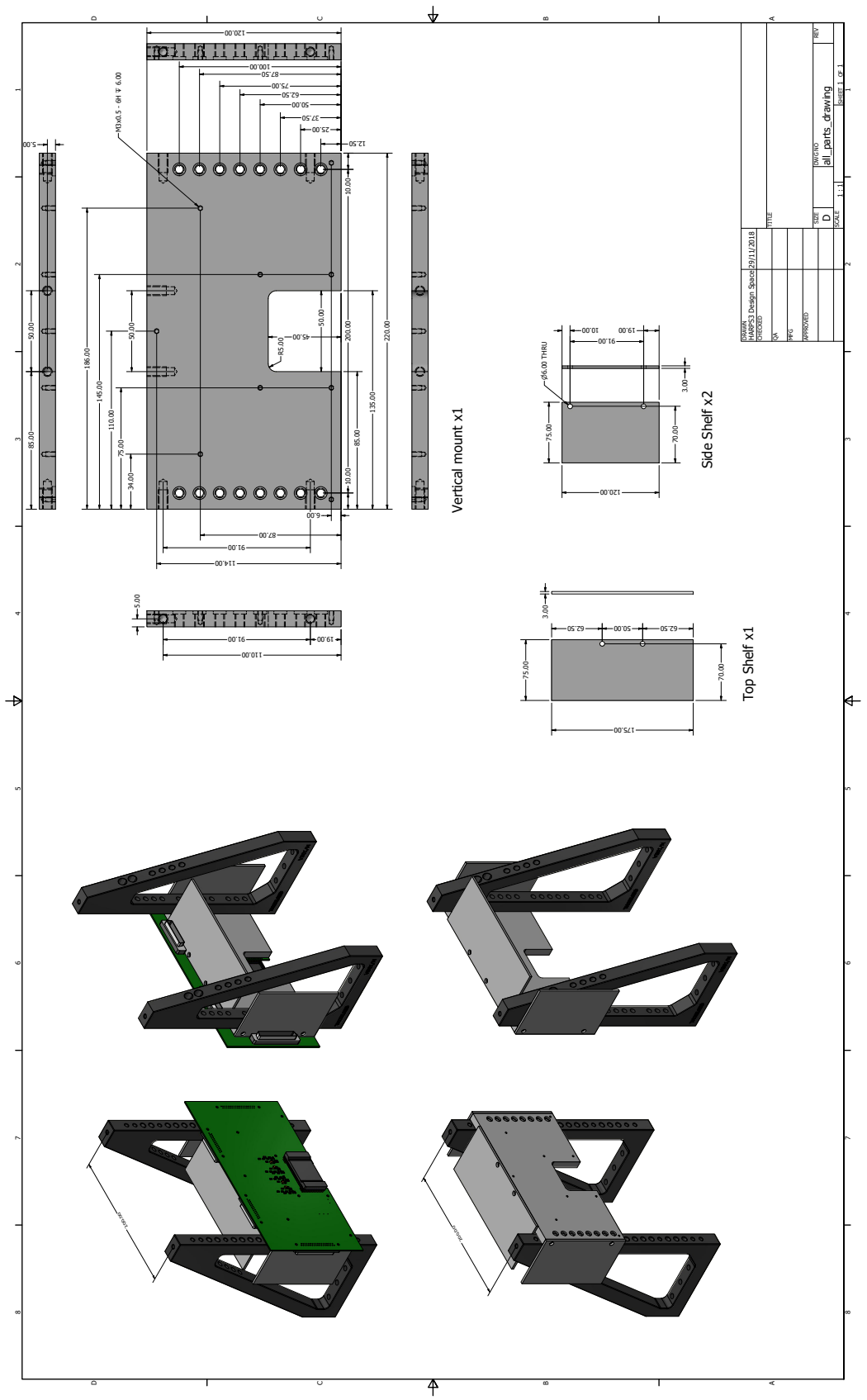


Figure 6.6: The CAD drawing for the PCB mount that incorporates the brackets for the optics bench. The side and top panels are used to support the weight of large cables that communicate between the PCB, and the CCD controller and pre-amp.

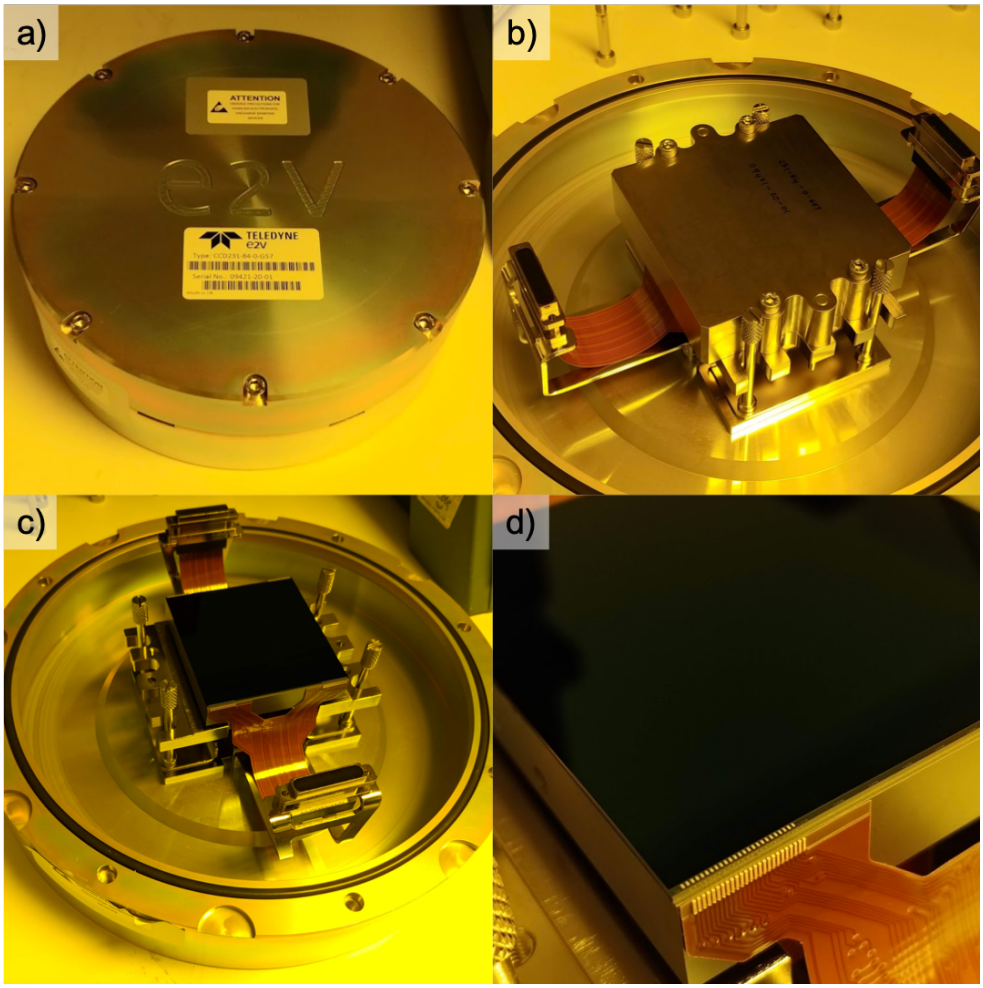


Figure 6.7: The inspection of the HARPS3 science CCD. In a) the large cylindrical enclosure protects the entire assembly both physically and environmentally. In b) the black rubber o-ring around the edge is used to create the seal, and the electrical connections to the CCD are protruding from each side. In c) the metal protective case is removed revealing the CCD surface. In d) the delicate gold bonds between the CCD and the connecting wires are just visible. For scale, the CCD is approximately 65 mm in side length.

accumulate. One notable feature of the CCD surface is actually how dark and *featureless* it is, this is partly due to the anti-reflective coatings that promote quantum efficiency.

6.4 Mitigating Environmental Effects

The variation of room temperature proved to be a limiting factor in fringe stability for the preliminary experiment of Section 6.1. Hence additional measures were taken to mitigate this.

Firstly, the entire experiment was moved to an ISO 7 (class 10,000) cleanroom. This cleanroom has a dedicated thermal control loop to monitor and maintain the temperature

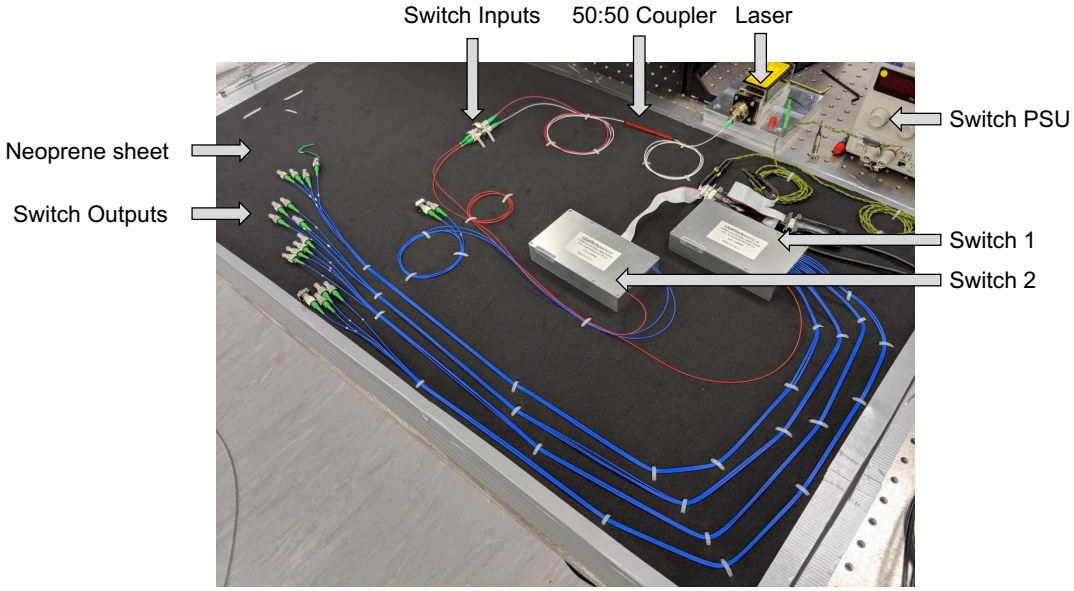
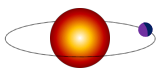


Figure 6.8: The fibers mounted to a Neoprene base to mitigate high frequency vibrations.

set-point to within ±1 K over a 24-hour period, and the humidity set-point to within ±1%. Additionally, the presence of dust and particulates in the air would be catastrophic for an open detector. The black-out box was designed with this in mind and is made from a non-shedding card-like material, but placing it in a cleanroom also will minimise the chances of a stray piece of dust landing on the detector. Entering the cleanroom required the use of overshoes, clean-room compatible jackets, gloves, face-mask, and a hair net to reduce the amount of particulates in the air.

The large optics bench is mounted on four large rubber feet to dampen any low frequency vibrations that may arise from the floor of the building. On the bench, the entire optical path is mounted to a 10 mm thick neoprene sheet to decouple any high frequencies vibrations on the table being transmitted to the fibers. Figure 6.8 shows the fibers mounted to this sheet and how they are pinned down to it. The fibers are gently pinned with clippings of plastic cable ties such that they do not bend, but resist any motion. The open fibers (16 from one switch, 2 from another) are later connected to other fibers that route to the fiber block.

Before the fibers were mounted as shown in Figure 6.8, the optical fringes were witnessed to randomly change in phase over a fraction of second. The chaotic and unpredictable thermal fluctuations and vibrations were causing small path differences between the fibers and affecting the phase stability. After mounting the entire optical chain onto the sheet, the fringes were stable over many tens of seconds, many times more than the chosen exposure time.



6.4.1 Monitoring the Environment

To track the temperature variations for possible correlation with any unwanted fringe drifting, a 8-channel temperature probe with ±10 mK precision was installed. Each of the thermocouple probes could be mounted to various components such as the fiber block, the CCD mount, or the optics bench. The module also has an internal cold-junction for reference, such that relative changes can be easily tracked.

The 8 probes are fixed to the metal surface of the optics bench, Figure 6.9, and the values are plotted. A typical 120-hour window of measurements is shown in Figure 6.10. The temperature variations are much larger than the expected ±1 K over a 24-hour period and include some very sharp drops. As mentioned in Section 6.2, we require at most ±1 K variation during a typical data collection run of a few hours. There are some smooth, flatter windows of around 8 to 12 hours that could be used for consistent data collection. Looking at a shorter window of time in Figure 6.11, the temperature of the clean-room was observed to oscillate over a period of around 20 minutes by upwards of 1 K. These oscillations tended to appear and disappear with no obvious source and are suspected to be a result of how the thermostat in the air handling system decides to either heat or cool the room back to the set-point. There was no correlation of these oscillations with clean-room activity or time of day, and they were mostly present at all times. These short term oscillations are not desirable for stable interferometric purposes as they will cause drifting fringe phase as the fiber baselines expand and contract, and they can change the periodicity as the path-lengths change. The data analysis procedure assumes that the fringe pattern is identical between frames and that only the phase varies. The difference between the cold-junction and the 8 temperature probes are likely due to the cold-junction being inside the temperature probe station whereas the probes were exposed on the surface of the optics bench.

A CCD is read out in series from one pixel to the next. For the 2048 × 2048 HARPS3 test CCD, even if it is clocked at 1 MHz it would still take 4 seconds for a full frame to be recorded. The exposure can be much shorter than this, but the exposure of the pixel is not taken *until* that pixel is next in line to be clocked. Thus, if the fringes were to unpredictably drift during this window, the spatial fit and data analysis would likely fail. Once the short period temperature shifts were observed, an optical shutter was installed directly in front of the fibers to allow a controlled burst of photons to reach the detector but keep it in the dark for the rest of the readout. If the intensity of the pulse is sufficient, each pixel will then contain the useful information of the fringe intensity plus some low level dark current for the duration of the read out.

The ESO controller used to drive the CCD also conveniently has a 5 V output specifically for driving the shutter of an optical telescope. Because the shutter was placed directly in front of the optical fibers, ∼1.5 m from the detector, a shutter-map was not necessary as the fibers are

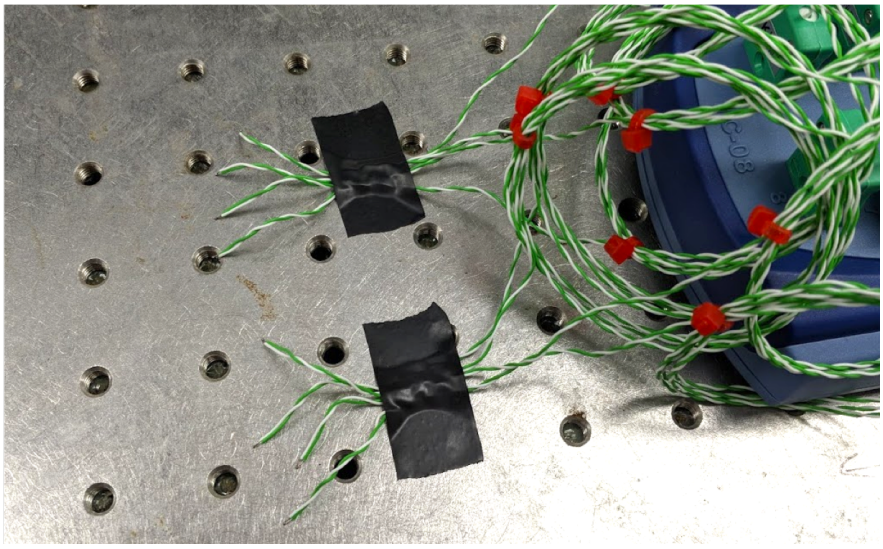


Figure 6.9: The configuration for the temperature probes during the environmental monitoring stages. This set up was placed in the middle of the optics bench and off to one side such that it was not disturbed over many days.

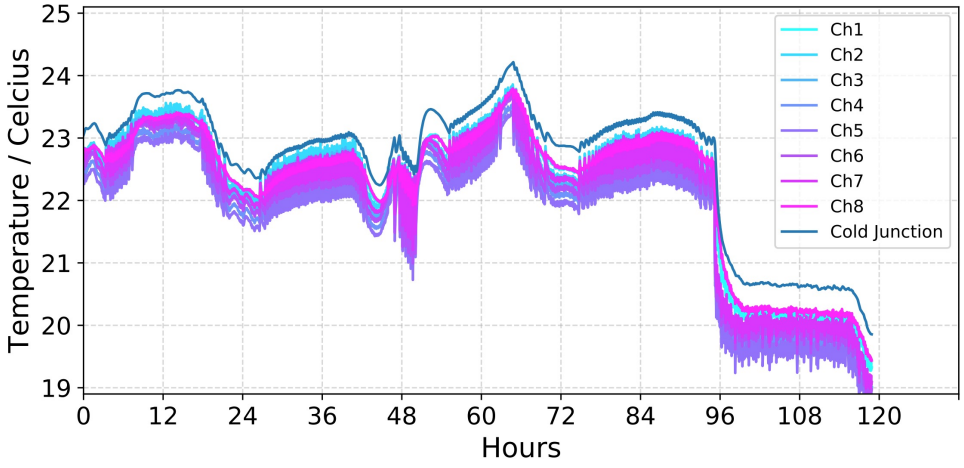


Figure 6.10: A 120-hour section of the clean-room temperature across the 8 probes all mounted to the optic bench. The variation is much greater than the anticipated ±1 K stability over a 24 hours range. The window is chosen to display the different types of variation but is typical for any given 120 hour window. The sudden drop in temperature at 96 hours is typical of how unpredictable the air-handling system was.

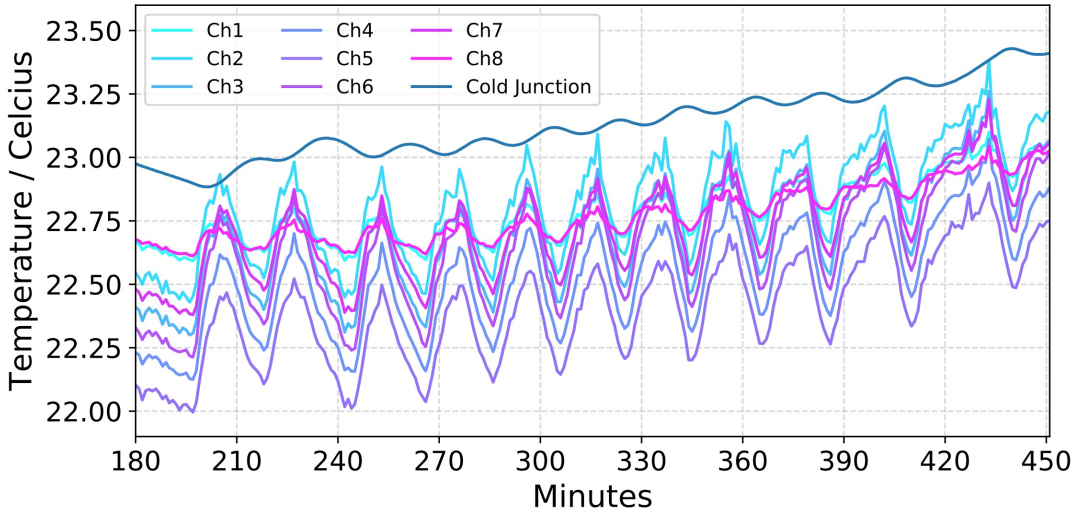


Figure 6.11: A zoom of Figure 6.10 showing short period temperature oscillations of around 1 K over a 20 minute period.

a near point source that were aligned with the centre of the shutter diaphragm.

6.4.2 Consideration of Air Turbulence

If there is substantial movement of air along the beam axis between the CCD and the fiber tips, the fringes could be non-uniformly distorted. This would cause large deviations in the spatial fit which would manifest as pixel displacements.

There should be minimal air motion inside the blackout enclosure as the box is completely sealed and there are no sources of heat inside that could act as a source of rising hot air which would lead to turbulence. All of the major sources of heat, such as the laser module and power supplies, are placed outside on the optics bench. The large clean-room air handling unit supplies air through a number of vents located near the floor in the corners of the room and the optics bench is situated in the middle so most of the air motion stemming from the supply is likely reduced by the time it reaches the bench.

No further measures were taken to mitigate the motion of air, but future plans were considered should the need arise. These included a large tube to cover the beam with a non-reflective surface on the inside to 'protect' the propagating fringes, and cutting a series of small holes in the blackout box to allow the temperature to regulate. Alternatively, the entire clean-room is a completely isolated 'room inside a room' with no natural source of light. It would be possible to actually completely remove the blackout enclosure and simply close the door. Whilst this would leave the delicate CCD and fiber tips exposed to the open air, it might be an effective method of reducing any air turbulence in the enclosure.

6.5 Data Collection Procedure

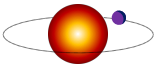
Following similar procedures of Shaklan et al. (1995) and Crouzier et al. (2012), and the results of the simulations in Chapter 5, the procedure for obtaining a dataset was as follows.

1) Decide baseline separation and angle, set up optics appropriately
2) Record the optical parameters and measurement errors of the baselines z, s, θ, for the spatial fit
3) Collect dark and flat frames for image processing
4) Acquire desired number of frames (at least 100) during a 2π phase sweep
5) Rotate block by 90° and repeat from 3).
6) Change baseline s, and/or baseline starting angle θ, and repeat from 3) for a new dataset.

Then, with all the data collected, the procedure for analysing and generating pixel maps is as follows:

1) Apply flat-field and dark current corrections to all images
2) Complete a spatial fit of all data stacks with the measured geometric parameters and errors to be used as initial guesses and bounds for the fit
3) Output the fit stack with a sensible file-name for per-pixel analysis
4) Loop through two pairs of stacks (a given baseline, the fit, same baseline with 90° rotation) and compute the pixel displacements as per Sections 5.1.5 to 5.2.
5) Output a pixel displacement map for a given baseline
6) Repeat for all data stacks of a given detector

During the spatial fit, it is possible that a particular frame results in a bad fit as the detector may have an erroneous readout, the background may vary, or other issue may arise. Hence, a particular frame index can be flagged if it failed, or if a critical parameter that defines the period (z, s, λ) lay outside some predetermined range, such as a certain threshold of standard deviations of the entire set. Additionally, during the temporal fit, a similar filter can be performed for any data point (a pixel brightness at a certain time-stamp) which can be removed if it lies out of a median filter around the sinusoidal signature. This particular data-point will also indicate an entire frame may be erroneous which can be removed. For my initial investigations I only rejected the frames if the fit was reported to fail completely but the functionality for some more advanced data filters were partially investigated but not implemented. Each generated pixel displacement map for a given baseline and angle can be averaged to obtain an estimate of the true position of the pixel positions with some uncertainty.



6.6 Testing the Experiment with the CMOS Sensor

The pixel mapping procedure was performed on the Thorlabs CMOS camera as a test of the stability of the optical components. As mentioned, the pixel structure of a CMOS sensor is inherently different from that of a CCD. For each pixel, only a quarter of the surface area is used to collect photons, the rest of the space is for a per-pixel amplifier circuit. Each pixel is therefore only, at maximum, 25% efficient. However, CMOS sensors can be read out many times faster than a traditional CCD sensor so this camera can be used to assess any short term fringe variability.

The fibers were spaced at $z = 2.5\,\text{mm}$ to create the largest optical fringes possible, and the aim was to collect 100 phase steps at an angle of 45°. However, during this procedure it became apparent that the phase modulator did not function as intended and acted to destroy the optical coherence and, hence, the fringes. The piezo crystal is driven by a high-voltage supply between $-200\,\text{V}$ to $200\,\text{V}$ which linearly stretches the crystal to add extra path length. For the wavelength of the laser used, this corresponds to maximum possible phase shift of $\sim -2\pi$ to 2π. The intention was to sequentially apply a gradually increasing voltage and to automatically capture frames at each step. Because the high-voltage supply can handle increments of just a few mV it would have been possible to capture many hundreds of images for a complete cycle. However, when the voltage was set to even just low values of this range, the fringes would lose contrast, gain additional interference structure, and eventually disappear at just over $100\,\text{V}$. Figure 6.12 displays the results from completely bypassing the phase-modulator, i.e. static fringes, through to routing the beam through this optical set up and increasing the voltage. It is clear here that the expanding crystal is causing fringe decoherence and also introducing additional interference at various angles.

It was noted that the fringes *did* sweep when the voltage was applied and so the unit did function as an optical phase modulator, but not enough to obtain a full period before the decoherence would be high enough to cause the spatial fit to fail. It was suspected that the piezo-crystal was deforming in a non-uniform pattern which would slowly destroy the coherence of the beam emanating from the front surface. It was not possible to test this hypothesis as the phase-modulator only has a 2 mm aperture to ensure good optical alignment with the beam axis.

The phase-modulator was removed from the optical path (see Figure 6.13) and a investigation into a suitable replacement was started. Because the clean-room and the optical bench were subject to slow temperature variations on the scale of 20 minutes, see Figure 6.11, the fringes would naturally drift on these timescales as the components would expand and contract along the beam axis. Provided that the actual baseline control of the block did *not* expand past some

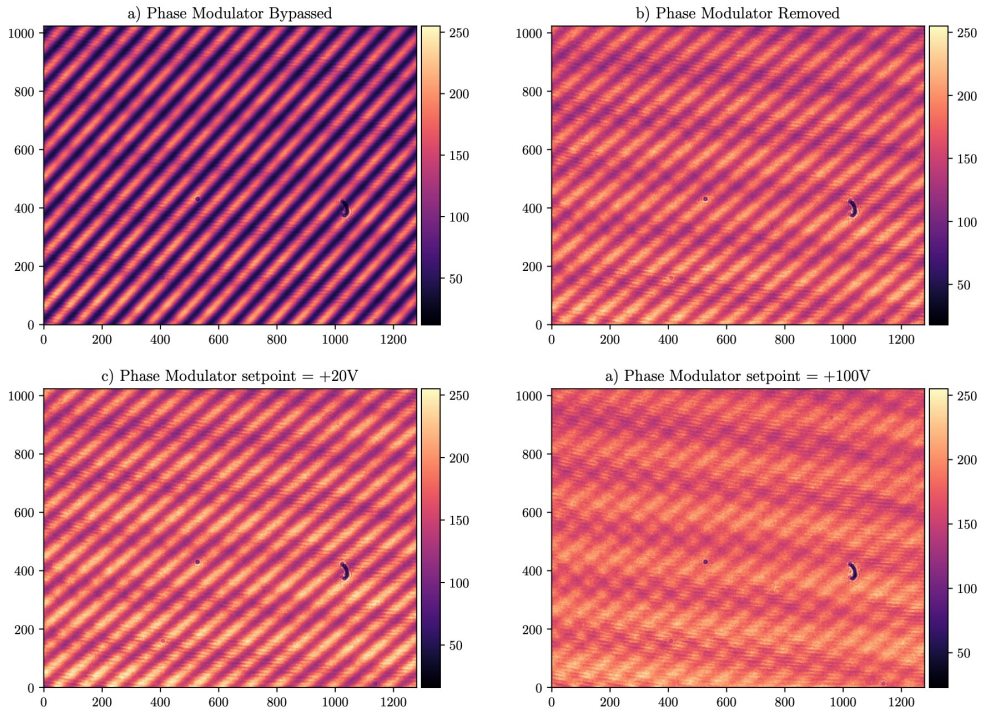


Figure 6.12: The effects of the piezo-electric phase-modulator when applied to the optical fringes. In a) the phase modulator was completely bypassed and the light was all in-fiber. In b) the beam was launched out of and into a fiber with no phase modulator present. In c) and d) the phase modulator was installed in the open beam and a fixed voltage applied. The black object towards the right of each frame is a piece of debris.

measurable threshold, the natural oscillations could be used as a mechanism for sweeping the phase. Hence, step 4) in the data collection procedure was replaced with 'Capture many frames and allow the natural temperature variations to passively sweep the phase'. Then, an additional step of the data analysis procedure is placed before number 3) of 'sort the data as per the fitted ϕ of each stack'. Provided the full phase range is covered, the single period sine-wave of the temporal fit will not be affected.

To test the pixel mapping procedure with the CMOS sensor, the optics were set up such that $z = 1.645\,\text{m}$, $s = 2.5\,\text{mm}$, and $\theta = 45°$ which creates fringes as seen in the first frame of Figure 6.12. 500 frames were captured at a rate of around 10 Hz and the block was rotated by 90° and 500 more frames recaptured. 500 frames were chosen via manual testing to ensure a full phase sweep of the fringes on the detector. Because the analysis is much faster when the CCD is of size 2^N, the 1024×1280 pixel image was cropped to the central 1024×1024 pixel section.

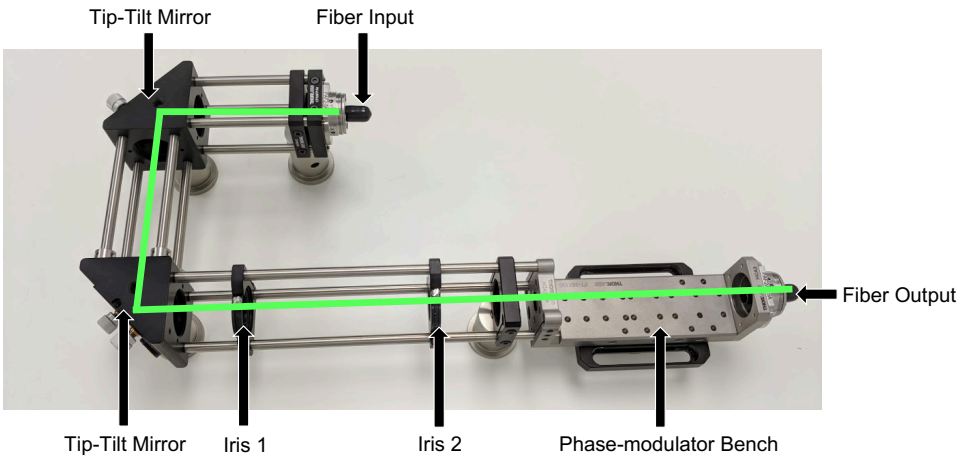
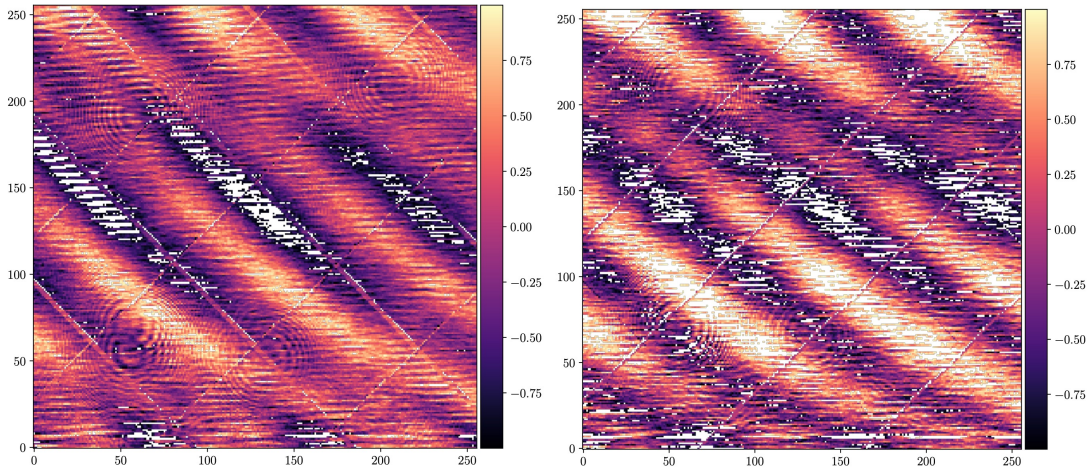


Figure 6.13: The U-bench used to launch and align the open beam through the phase-modulator and re-couple back into a single-mode fiber. The green line is the path followed by the beam.



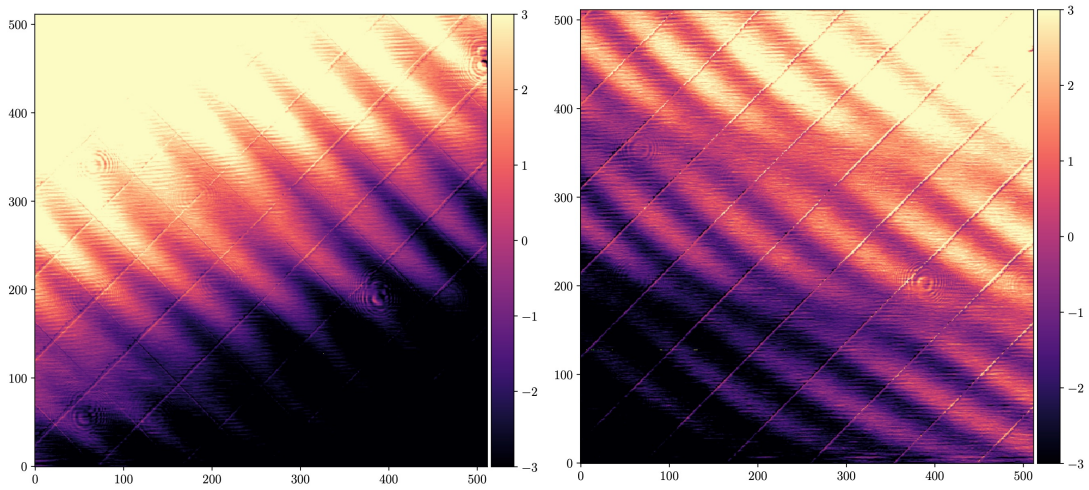
(a) X Displacements

(b) Y Displacements

Figure 6.14: a) The X displacements for the CMOS sensor cropped to a 256×256 pixel region, b) the Y displacements. The scale has been limited to a ± 1 pixel range, anything greater has been coloured white. The pixel maps exhibit many systematic patterns that are discussed in Section 6.7.

6.7 CMOS Pixel Map Results

The preliminary results from the CMOS sensor are shown in Figure 6.14. These results exhibit a number of interesting features that are either systematics of the analysis procedure or the result of environmental instability, and as such may not be a true reflection of the displacements. A larger crop of the pixel map is shown in Figure 6.15, here I have selected a 512×512 pixel window and allowed the colour range to extend to ± 3 pixels to show the overall structure of the



(a) X Displacements

(b) Y Displacements

Figure 6.15: a) The X displacements for the CMOS sensor cropped to a 512×512 pixel region, b) the Y displacements. The scale has been increased to a ± 3 pixel range. The x, y units are in pixels.

measured displacements despite being non-physical results.

The full displacements distribution is shown in Figure 6.16. Here there are many displacements greater than 1 pixel, with a small group of measured displacements lying at ± 4 pixels. These outliers are likely a product of the temporal fit generating two values that are only different by, say, 0.1π rad but their true phase values are 1.0π rad and 1.1π rad. Because the temporal fit is limited to $\phi = \pm 1\pi$, the later gets fitted as -0.9π rad instead and hence the phase difference is measured as 1.9π rad. This then manifests as a large number of pixels when deprojected. To catch these outliers, I included some statements in the pixel deprojection loop that would identify when the measured phase difference exceeded 1.0π rad, and would appropriately add or subtract 2.0π as necessary to the correct value. However, as seen in Figure 6.16 there are still a small number of anomalous results which manifest as large pixel displacements.

In Figures 6.15a and 6.15b there is a clear cross-hatched pattern angled at 45° which is the same as the fiber angles. This could a result of the spatial fit not quite finding the true angle of the fibers which would then leave residuals in each frame between the fit and the data. Secondly, there are features that resemble the concentric rings of an airy disk in both images at similar locations which are likely artefacts of small pieces of debris on the detector causing residuals between the data and fit which propagate through as displacements. There are also a number of patches of pixels whose displacements were measured to be greater than 1 pixel which is obviously non-physical. This indicates a large discrepancy between the expected and measured intensity variations as a function of time for these pixels. This could be attributed to

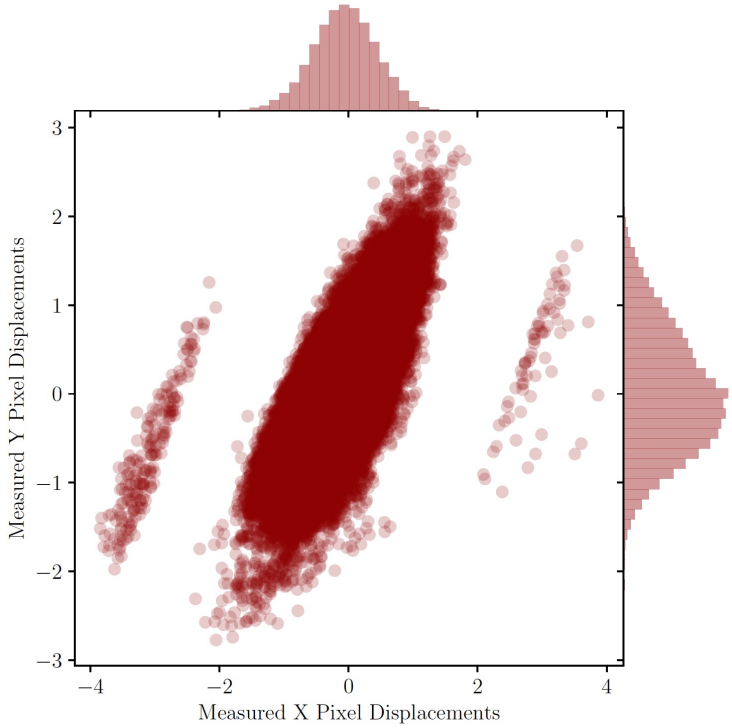


Figure 6.16: The bivariate distribution of the measured displacements of the CMOS sensor.

the temperature fluctuations measured in Section 6.4.1 which would cause some or all of the spatial fit parameters to vary as a function of time, instead of just the phase. If this is the case, the relative drift between the fibers and the CCD in either x, y, or z would render the analysis inept as the successive frames are essentially a new baseline and can not be analysed as a set.

To investigate this, I have plotted the fitted parameters for each frame in Figure 6.17. Despite the relatively tight bounds, the poor image quality from the CMOS sensor is causing the fit to vary between frames which effectively nulls the pixel displacement measurements as the analysis assumes consistent fringes. Two of the parameters are well constrained by the data: the fiber angle θ and the fiber core diameter ω. This is expected as a slight change in angle would cause a large difference in the least-squares, and the core diameter controls the overall sinc envelope only.

In an ideal world, the bounds of the parameters would be fairly unconstrained so as to not introduce any bias into the spatial data. However, leaving the solution completely unconstrained would lead to potentially erroneous solutions, local minima of the least-squares, or an unrealistic convergence time.

From Figure 6.17, the angle θ is clearly preferred at the lower end of the bound which could indicate the true angle is not what I assumed, which for this iteration was $45 \pm 1^\circ$ assuming

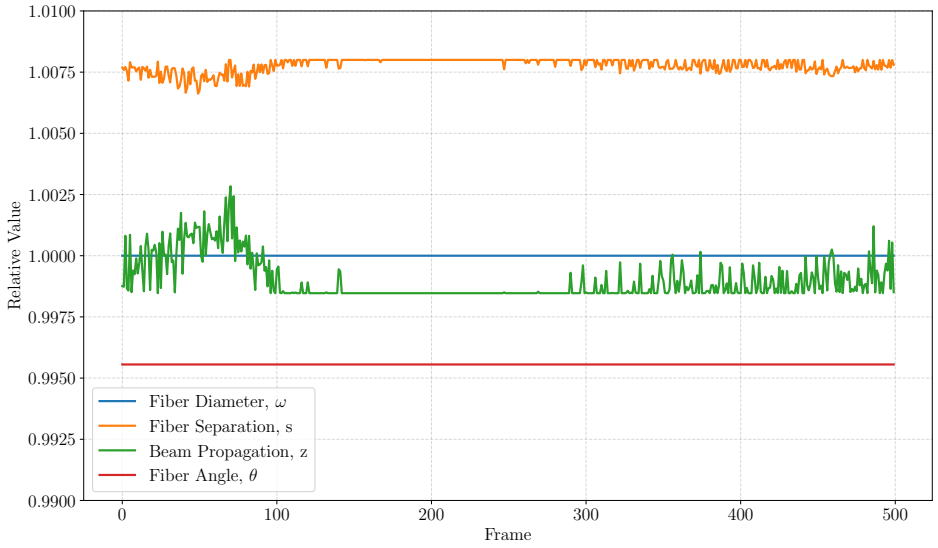


Figure 6.17: The fitted parameters as a function of frame number (time) for the CMOS sensor normalised to the measured value of that parameter such that relative variations can be assessed. For the fiber diameter, ω, the literature value of 4 µm was used.

the sensor was mounted precisely in its enclosure. To rectify this for future analysis, I instead calculate the fringe angle by performing a 1D Fourier transform of all of the rows of the spatial data and all of the columns to obtain the fringe periodicity in the x and y axis. I average all of the rows and columns respectfully and then compute the 2D wavevector from this of the entire frame. I then use this value as the first guess of each spatial frame and give the bounds as a 1° error.

The spatial fit also favours any first guess I give for the fiber diameter ω. This is due to the overall structure of the pattern depends the least on the fiber core diameter as it only constrains the width of the sinc envelope. As this experiment is designed such that the CCD samples the central lobe with relative flatness, this parameter could likely be omitted to speed up the parameter estimation.

Also from Figure 6.17, the fitted values of fiber spacing s, and beam propagation z vary at the start and at the end of the data collection period. The relative flatness from frame 150 to 300 is when the fit has chosen one of the bounds as the best guess for each successive iteration. These parameters have an inverse affect on the fringe periodicity so it is not unexpected that as one increases the other decreases to compensate. But is not obvious what is causing this deviation as a 0.25% change in the z propagation distance corresponds to a 4 mm variation in physical distance which is not physically possible given the conditions of the lab.

Due to time constraints and the imminent installation and preparation of the HARPS3 Test CCD, no further data was collected on the CMOS sensor. This was largely justified with the fact that the analysis was designed upwards specifically for a CCD sensor and not a CMOS sensor, and any investigation into the specifics of the pixel displacements measured could be simply due to the low quality of the CMOS sensor.

6.8 HARPS3 Test CCD Calibration

In this section I detail the procedure of collection and analysing the fringe data for the 2048×2048 pixel CCD. I then discuss the results of the pixel displacement maps generated from the pipeline.

6.8.1 Installing the HARPS3 Test CCD

Installing the HARPS3 Test CCD was a lengthy process that involved the construction of the server rack to house the CCD controller power supply and the Linux machine that drives the controller, see Figure 6.18. The custom CCD PCB and mount were designed and manufactured well in advance as the e2v chip model was known early on in the experimental process.

The PCB was tested without the CCD mounted to ensure that the ESO controller was sending correct clocking signals and voltages to the various CCD channels. I give my personal thanks to Dr. Bodie Seneta for giving many months of his time during this project to ensure the entire CCD side of the experiment worked flawlessly. Without his help here the experiment would have been limited to the CMOS sensor.

One very large project completed by Dr. Seneta was in writing the configuration file for the CCD as there were no obvious documented versions that were compatible with this CCD. This involved a lengthy testing process of trialling various configuration options such as clock speeds, gain settings, pin voltages, until the detector was able to produce images of optical fringes.

6.8.2 Dark and Flat-field Frames

For each set of data (a single baseline at a given angle), I collected a set of flat-field, e.g. Figure 6.19, and dark frames. As a single set of data takes around 1 hour to collect, it was assumed that the flats and darks are applicable to this set only. When the baseline is rotated by 90° or a new baseline is chosen, new flats and darks are collected. An example of a dark-frame is shown in Figure 6.20.

For the dark frames, I turned off all sources of light within the cleanroom and I collected 100 frames of a 10 second exposure. These frames are then averaged into a ‘master dark’ frame and analysed for the average contribution of dark current present over 10s.

Figure 6.18: The experimental set-up with the large server rack placed behind the optics table that holds the blackout enclosure.

For the flat-field frames I followed the procedure of Wheatley et al. (2018) by using an Electro-luminescent Panel (ELP) to uniformly illuminate the detector. I mounted the panel behind a diffuse polymer screen to further decohere the light from the panel. The panel was placed approximately 20 cm from the front surface of the CCD, and a neutral density filter placed directly in front of the CCD to lower the counts to a sensible range as the panel would saturate the detector at the shortest exposure time possible.

The ELP has the colour of 'Tron Blue' with a wavelength of approximately 475 nm which enables the Blue Diamond effect to also be mitigated. The blue diamond effect arises from non-uniform surface profiles created during the back-thinning of modern CCDs. This spatially dependent pattern is also wavelength dependent and does not appear at redder wavelengths (Wheatley et al. 2018). One potential downside to this technique is that we are not measuring the flat-field at the wavelength used for the fringe data.

The flat-field frames and dark frames allow for the characterisation of a few standard cosmetic defects often observed with CCDs. Dead pixels, hot pixels, and column defects are usually rare with science grade CCDs, but are more likely to occur with this engineering-grade

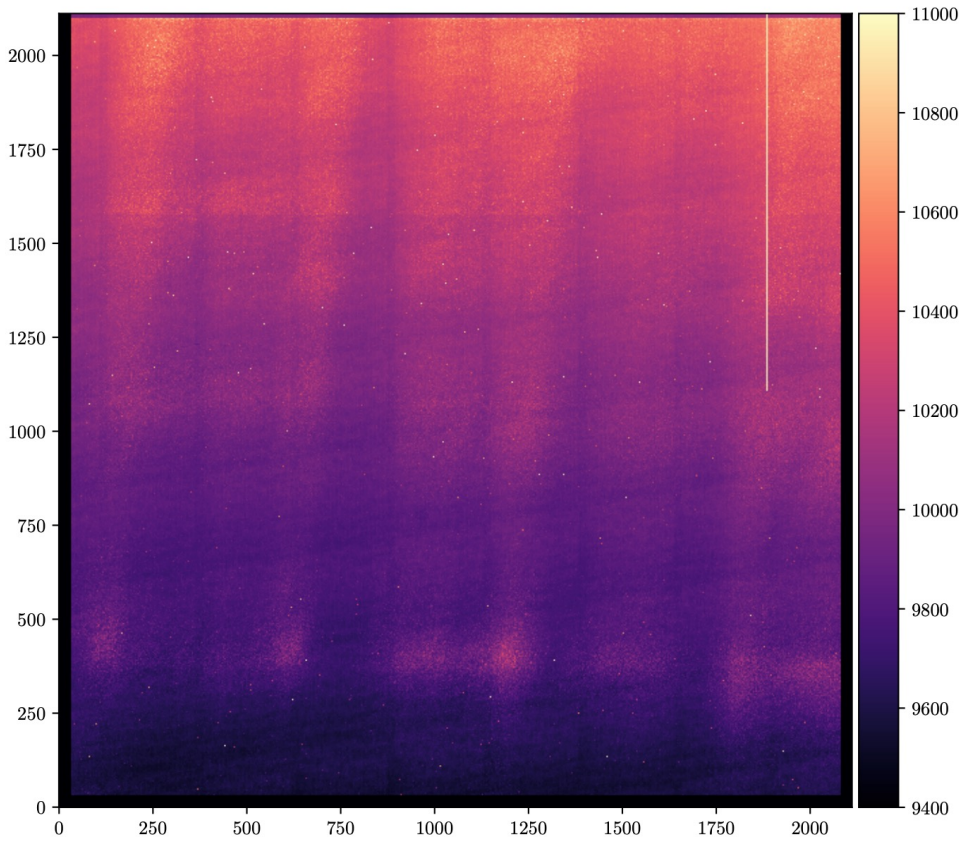


Figure 6.19: A typical flat-field frame for the 2048 × 2048 pixel CCD. A number of cosmetic defects are visible such as hot pixels, dead columns, the 'blue diamond' effect, and a global gradient from top to bottom.

detector. A large number of dead or hot pixels could cause the spatial fitting of the fringes to be inaccurate so a map of bad-pixels is created from the flat by conducting a rolling median filter of the average of many flat-field frames.

Figure 6.19 shows an example flat-field frame obtained from the ELP passing through an ND=2.0 filter to reduce the intensity. The image shows the CCD exhibits many common cosmetic defects such as hot pixels (the bright individual pixels), a dead column at the upper left corner, an overall gradient from top to bottom, the 'blue diamond' effect, and some smaller scale brighter regions that appear to not be randomly placed. The large number of hot pixels is due to the detector being of lower quality than a typical science-grade device.

6.8.3 Timing of Exposures

The readout rate for the CCD could be controlled between 50 kHz to 2000 kHz. Despite the read noise being highest at the upper end of the range, I chose a rate of 1000 kHz. This means

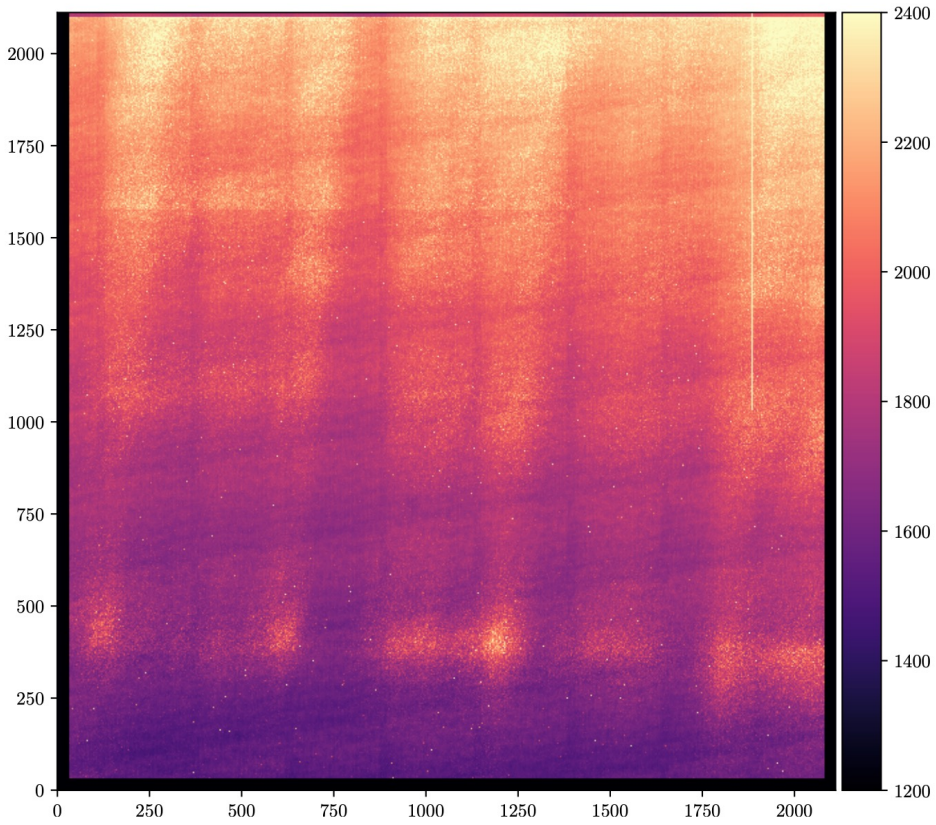


Figure 6.20: A typical dark frame from the 2048×2048 pixel CCD. The structure is very similar to the flat-field of Figure 6.19 which might indicate the presence of background blue light that is not visible to the eye, or some structure inherent to this particular detector. The same hot pixels and dead column can be seen which verifies that the features seen in the flat-field are illumination independent.

a full frame readout takes approximately 4 seconds. The CCD datasheet specified a read noise of just $1\sigma = 8$ electrons for a 1 MHz readout speed which is very low when compared to a pixel that can contain up to 150,000 counts at full saturation. The CCD controller allowed for multiple ports to be used, i.e. the CCD can be read-out in parallel across multiple channels to speed up the data collection. I elected to noise use this collection method as each channel would need calibrating for gain and noise variations, and the subsequent analysis scripts would need vast alterations.

However, during these 4 seconds the fringes have to be perfectly static for a frame to be spatially consistent. As the frame is being read out sequentially, if the fringes were to drift by even a fraction of a pixel, the spatial fit would be skewed by this drift and the displacements would be meaningless. Hence, an optical shutter was placed at the fiber tips and was connected

to the CCD controller. The CCD controller software can then clock the CCD with a controlled pulse that opens and closes the shutter to manage the exact amount of light measured by each pixel.

A typical exposure time for the optical fringes is around 0.1 seconds which resulted in roughly $\frac{2}{3}$ saturation of the pixels at the peak of the fringes. Thus, each pixel will register approximately 3.9 seconds of dark current and background light and 0.1 seconds of photons from the fringes. This is advantageous for fringes that are potentially subject to smearing on the detector due to temperature variations with the optical set up.

6.8.4 Calibrating the Switches

The switches contain 18 outputs in total for fine control of the baseline, and mounted to each is a single mode fiber with a male-male connector. As single mode fibers are notorious for difficult coupling due to their narrow cores, the outputs of the switches and fibers were connected to a precise fiber power meter to assess if any of the outputs should exhibits low power. These channels would then likely not be used as they would result in low fringe contrast.

The laser was directly mounted to both switches to minimise path losses, the power was varied across the entire range, and the output power was sequentially measured and is shown in Figure 6.21 for the 16 channel switch and Figure 6.22 for the 2 channel switch.

The 16 switch channels are seen to behave fairly linearly with input power, but there is a wide distribution of values at the upper end indicating that each channel has fairly large variations in losses. There also appears to be no correlation between switch position and final power. For maximum fringe contrast the outputs must have the same power, hence for the first iteration of data collection the switches with similar outputs were pair-matched to ensure this. The spread of power of all 16 channels is almost 100% which would result in fringes of just 50% amplitude if the two extremes were chosen as a pair. The 2 channel switch also exhibits similar linear performance but with just a ~20% power discrepancy between the outputs.

6.8.5 Aligning the CCD

The analysis script assumes that the CCD is in a perfect perpendicular plane to the beam axis, and that the CCD coordinates only vary on this plane in the x and y axis. Should the CCD be mounted such that it sits at a slight tilt or angle relative to this plane, the pixels will be sampling the interference pattern each with a z value that is slightly off from the expected grid and hence the spatial fit will exhibit structured residuals.

To prevent this, the CCD must be aligned to the optics bench with enough precision to eliminate a physical displacement from a misaligned CCD being interpreted as pixel displacements.

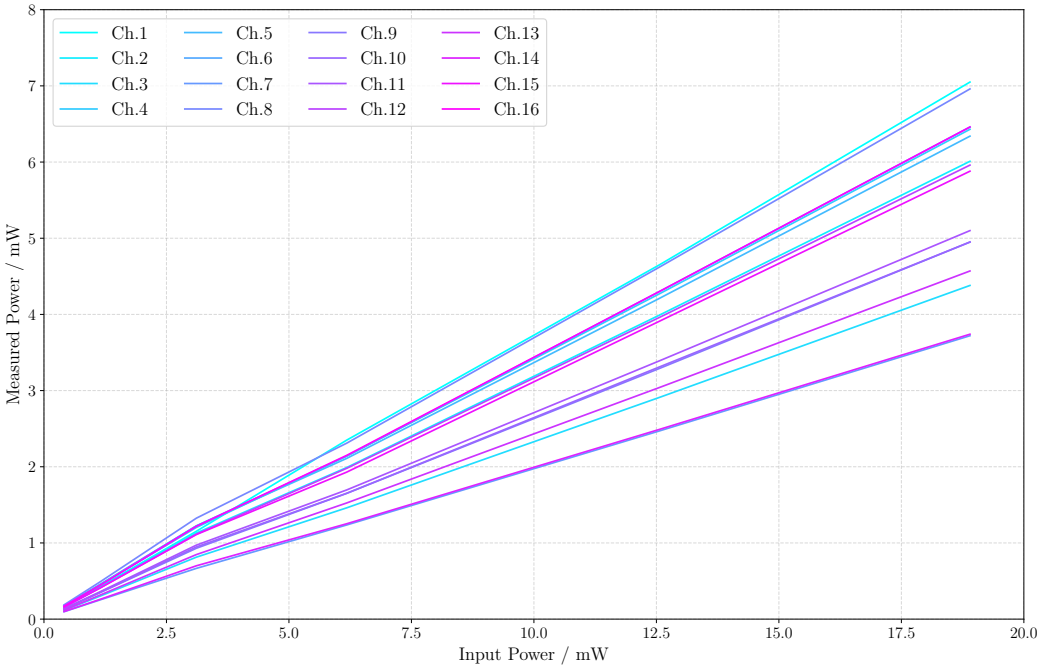


Figure 6.21: The power throughput of the 16 channel switch outputs where the input power is varied and the output is measured with a precise power meter. Each channel responds linearly but there are discrepancies between channels.

This process is especially critical for the 2048×2048 test CCD as this chip will be routinely removed from the mount during the initial testing phase of the experiment.

To measure the CCD alignment, a separate laser module is mounted in place of the optical fibers and a collimated pencil beam is aimed at the CCD. The reflection of the beam should return to the exit pupil of the laser if there is no misalignment. The laser can also be slightly translated in the x, y plane to check the CCD is evenly aligned across the whole surface rather than just a single spot measurement.

If the reflected 2 mm laser spot can be aligned by eye to within 1 mm, then the uncertainty of the angle of misalignment of the CCD respect to the optics bench draws an isosceles triangle of opposite side length equal to 1 mm and hypotenuses of $z = 1.5$ m. This angle is equal to just under 0.02°. The CCD mounting on the PCB board has an 'ever so slight' softness to it that allows for small adjustments whilst still maintaining good electrical contact. I used this feature to gently press on alternating corners of the CCD bracket within the mount to move the reflected laser spot into position. The CCD is then pivoted about its centre by this small angle and we consider the displacement of a pixel at the edge of the detector along the line of sight of the laser beam, see Figure 6.23. For a laser spot positional precision of 1 mm corresponding to a maximum misalignment of 0.019°, the edge pixel is at most displaced by 5.09 µm along the

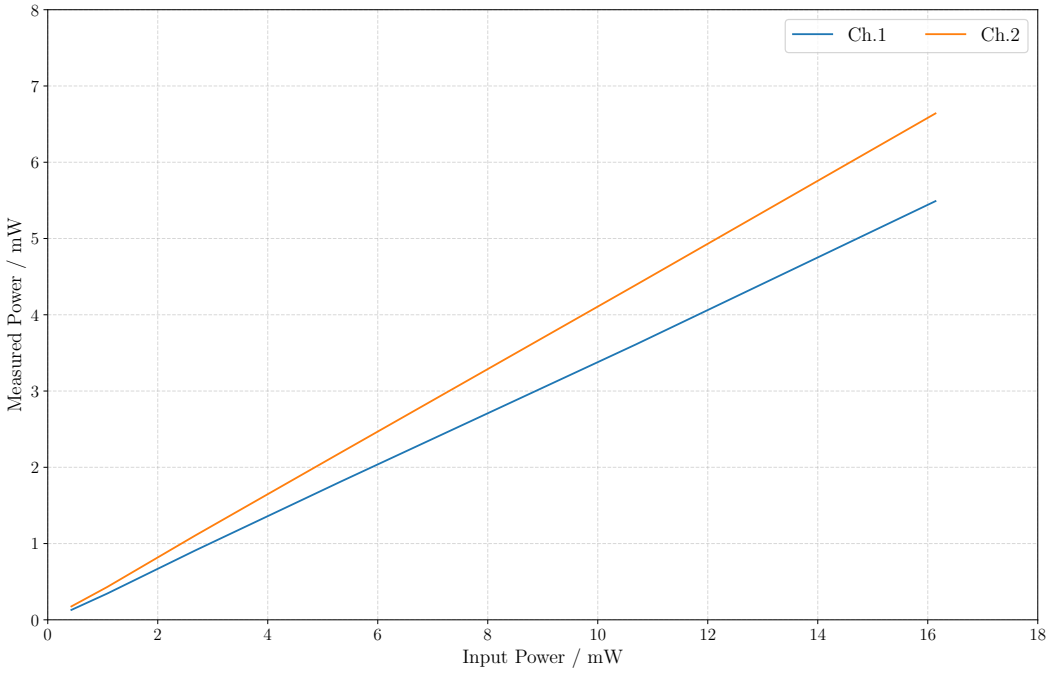


Figure 6.22: The power throughput of the 2 channel switch outputs where the input power is varied and the output is measured with a precise power meter. Each channel responds linearly but there are discrepancies between channels.

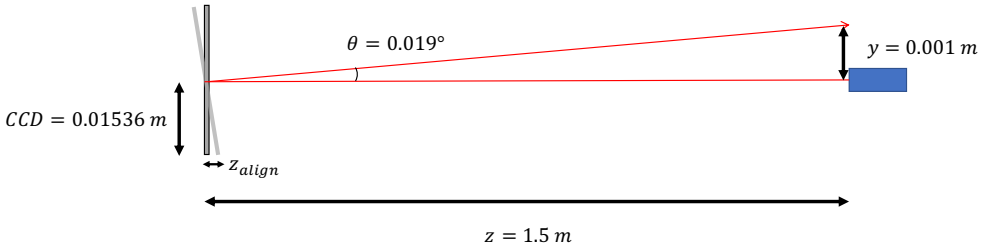


Figure 6.23: A schematic of the relevant quantities when aligning the CCD to the optics bench. The red laser propagates to the CCD and is reflected back, with a positional precision of 1 mm, the CCD is at most misaligned by 0.02°. The quantity z_{align} is the magnitude of the edge pixel's displacement along the z axis.

z axis, just $\frac{1}{3}$ of a pixel. This in turn results in a maximum x, y displacement of just 0.00017 pixels. In reality, the laser spot can be manually aligned to much more precision than 1 mm and as such CCD misalignment is not considered to be a major source of potential systematic error. It is therefore assumed that the CCD is sufficiently aligned to the optical beam axis when this criteria is met. However, if the CCD is not perfectly flat, the pixels will be sampling the interference pattern at a different propagation distance despite being well aligned to the axis. This may reveal itself when analysing the pixel displacements maps.

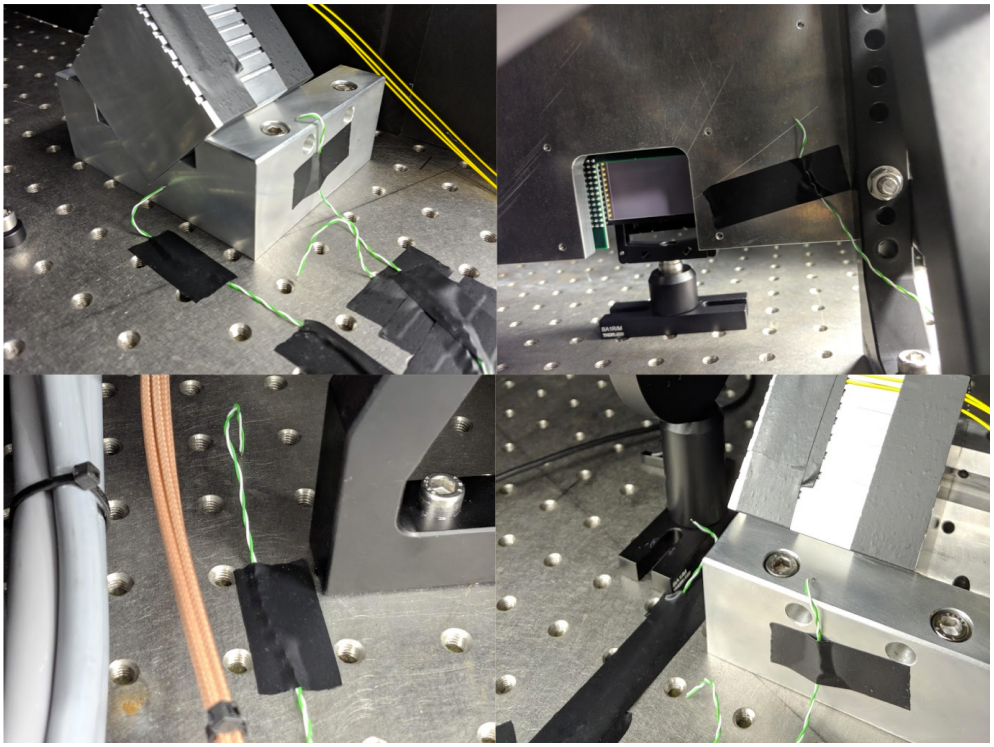


Figure 6.24: The positions of some of the temperature probes. Clockwise from top-left: 2 probes mounted to the fiber block and one on the optics bench, one probe mounted to the rear of the CCD mount, one probe in the open air by the shutter, one probe mounted to the optics bench behind the CCD mount.

6.8.6 Further Temperature Mitigation

After witnessing the wide temperature variations on the surface of the optics bench, Figure 6.10, the probes were then mounted to various components inside and outside the blackout enclosure to see how the observed air temperature changes affected the optical components, see Figure 6.24.

The probes were placed in the following locations: suspended in the air, mounted to the bench outside the blackout enclosure, mounted to the bench inside the enclosure, fixed to the CCD mount, mounted to the bench behind the CCD mount, fixed to the fiber mount, mounted to the bench by the fiber block. Thermocouple probes do not make good atmospheric temperature probes as they rely on a small surface area of contact which is impossible to achieve with a moving gas, but the probe in the air could make a useful reference.

The lab was left untouched over a 72-hour period and the data collected is presented in Figure 6.25 and a zoomed window is shown in Figure 6.26. Here it is clear that the measured temperature variations on the various surfaces and optical components are directly correlated with the air temperature variations caused by the air handling of the clean-room. The

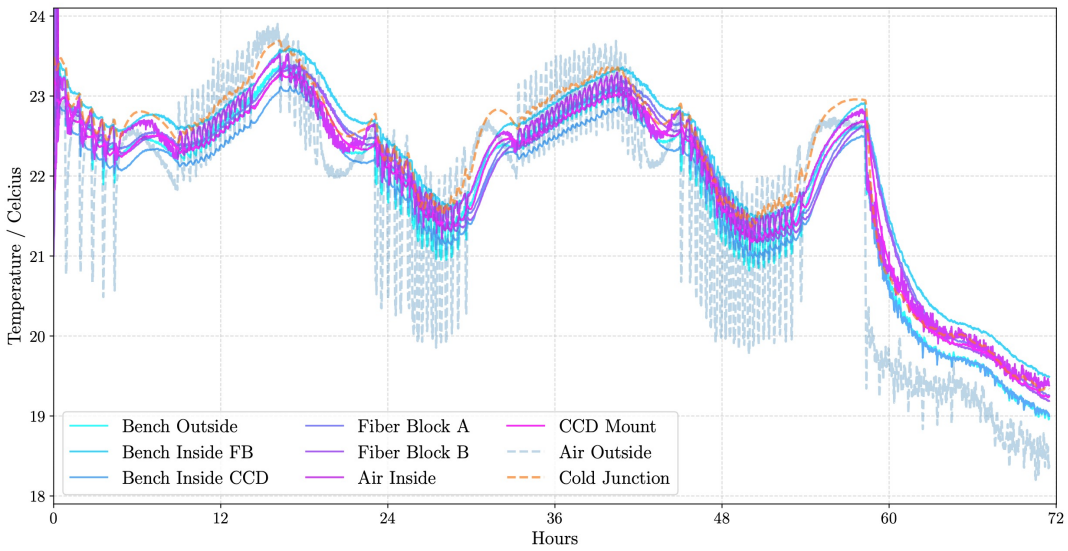


Figure 6.25: The temperature of the various parts of the optical set-up as measured over a typical 72 hour period. 'Inside' refers to a probe inside the blackout enclosure, and 'Outside' refers to one placed outside the enclosure, and 'FB' refers to the probe near the Fiber Block. The 'Air Outside' is a theromocouple probe suspended in the air above the optics bench outside the enclosure.

combination of sharp and gradual temperature variations cause small deviations in the baselines plus could cause air turbulence between the fiber tips and the CCD. These could account for the large pixel displacements as seen in Figure 6.14.

To further mitigate the affect of the air transferring the heat to/from the bench, 25 mm thick insulation was placed on all open metal surfaces to dampen the rate of transfer of heat from the atmosphere in the laboratory onto the optics bench and hence the components. The justification for this arose from examining the temperature variations of the cold junction that lies within the temperature probe enclosure. Here, the cold junction variations are damped when compared to those observed that are directly mounted in the optics bench. The hope was to emulate this damping but for the entire bench by reducing the surface area exposed to the air. I placed 25 mm thick polystyrene insulation boards on top of the blackout sheets to act as buffer between the temperature variations of the air outside and inside the enclosure, which is shown in Figure 6.27. Here the insulating sheets were cut to 'slot' inside the frame of the enclosure with the blackout boards placed behind. Where necessary, small holes were cutout to allow the fiber optics to enter unobstructed. Then, a further 144 hours of data was collected and is shown in Figures 6.28 and 6.29.

It is clear that the insulation of the blackout enclosure has not had a significant effect on mitigating the short-term temperature oscillations when they are present in the lab. It is

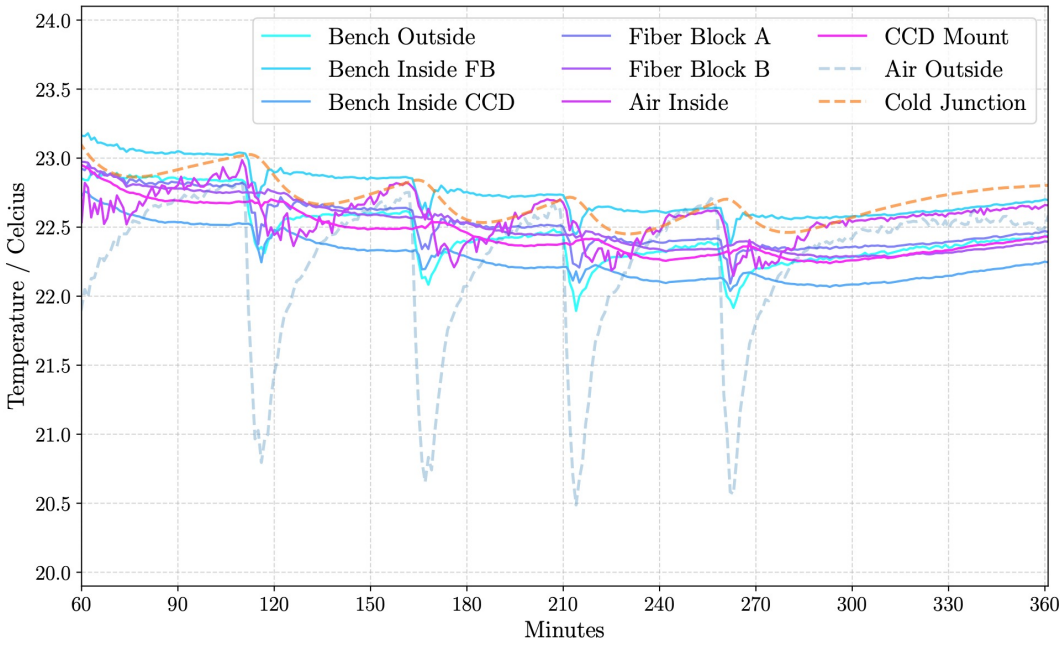


Figure 6.26: A zoom of Figure 6.25 highlighting the rapid changes of air temperature.

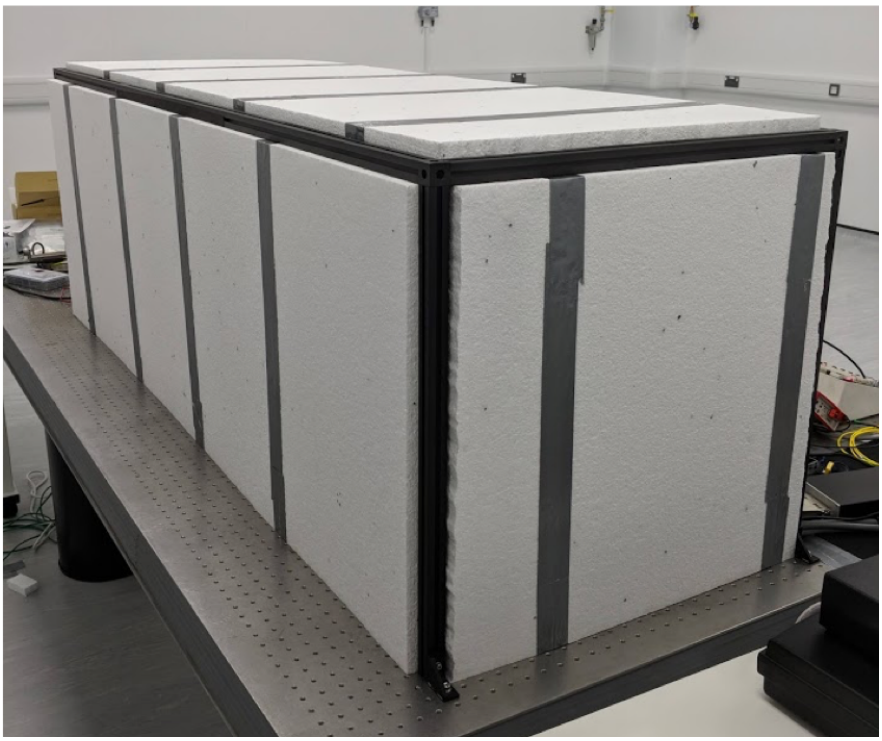


Figure 6.27: The insulation applied to the blackout enclosure to mitigate the temperature variations that affect the experiment.

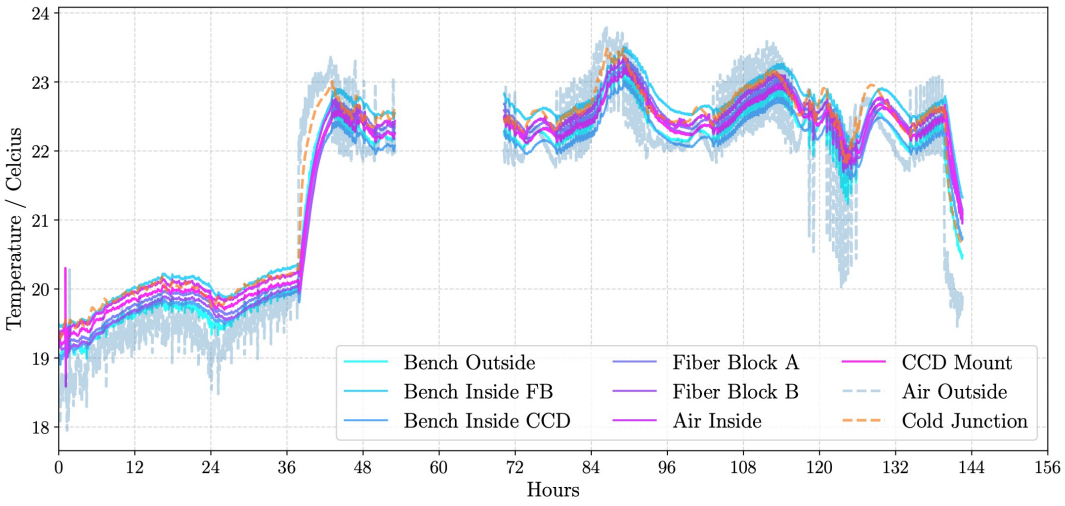


Figure 6.28: As in Figure 6.24 but with the insulation in place as shown in Figure 6.27. The spike at the start is my presence in the lab when installing the polystyrene insulation, and the gap at 52 to 72 hours is due to the lab computer unexpectedly shutting down so no data was collected here.

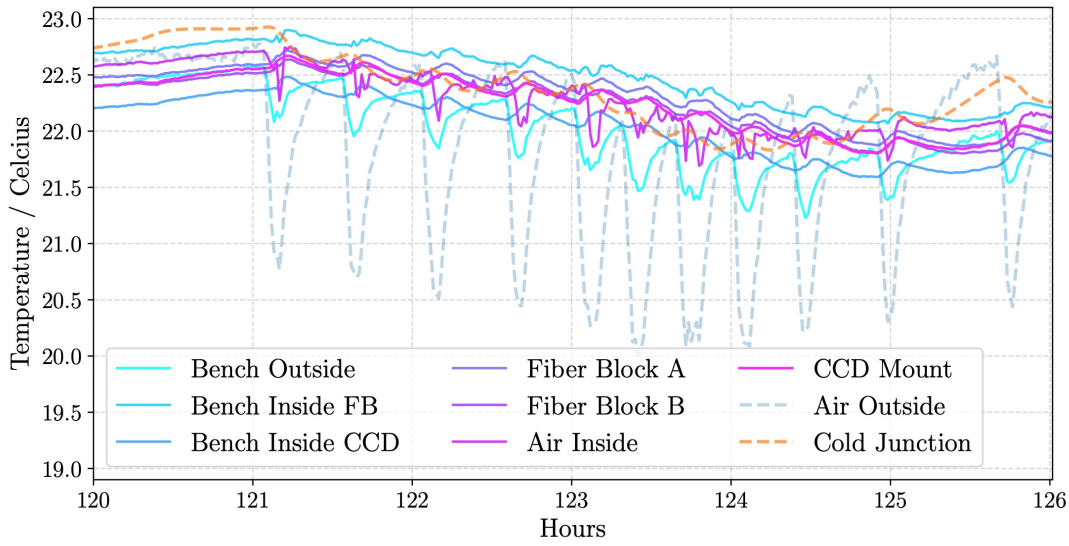


Figure 6.29: A zoom of Figure 6.28 showing a narrow range of 6 hours. Both the external and internal temperature oscillations are still present.

therefore likely that the large section of the optics bench that is exposed to air (both the top and underside) is acting as a heat sink which is causing the temperature variations observed.

6.9 HARPS3 CCD Results

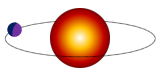
With the environmental variations mitigated as well as possible given the time constraints for this experiment, the fringe data was collected at once and then processed. In this section I detail the results of the processed data and outline what needs to be done to further improve the results.

6.9.1 Collection and Management of Data

The data was collected during a 2 month period in early 2019. Once a particular baseline was chosen with the optical switches, a set of dark and flat-field frames were collected immediately followed by the science frames of the fringes. As per the recommendations from the optical simulation, at least 100 fringe frames are required to handle the photon noise and to adequately sample the temporal domain of the fringe sweeping. Whilst the read-out takes 4 seconds, there are an additional few seconds taken by the controller to flush the data to the disk, wipe the CCD of charge information, and prepare for the next exposure. This process takes around 15 seconds in total which means taking 200 science frames plus the flat-field and dark frames is roughly an hour of data collection.

Once this process is complete, the fiber block is manually rotated by 90° and the data collection is repeated. An alternative strategy is to collect the data of a single orientation but of multiple baselines automatically by remotely changing the switch position. However when the block would then be rotated after this longer period, say 24 hours, the environmental conditions of the clean-room may have change sufficiently to render the two complimentary sets meaningless as the baselines may have changed between the two periods. Hence until the cleanroom was verified to be stable over longer time-scales, the former approach was taken.

A typical frame is approximately 10MB in size, so a full phase sweep of 200 frames with the 90° rotation is approximately 4 GB. Repeating this for n baselines starts to rapidly increase the disk space required to a significant fraction of TB. The local CCD controller computer only had an internal drive of 1TB which included the OS, relevant software, and many hundreds to thousands of test frames and calibration frames. Hence to mitigate this, once a fringe stack was completed the data was moved to the Cavendish Astrophysics cluster where the analysis scripts were hosted which also has daily backups for security. Then the local disk was flushed of the most recent set of data once the transfer was complete.

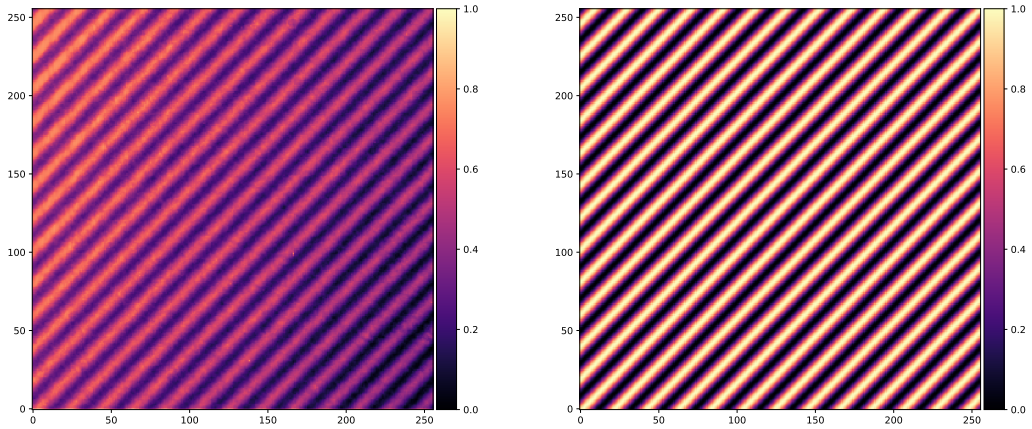


Data was collected for 2 of the shortest baselines (widest fringes on the detector) that correspond to a fiber separation of 2.5 mm and 3.5 mm. The data collection took a lot longer than expected due to many issues that arose. These included small amendments to the mounted positions of the fibers on the Neoprene mat, better optical insulation through the blackout box, re-mounting the fibers to the fiber block for better stability during the 90° rotation, mitigating reflections from the optics bench with non-reflecting optical tape, updating the automatic collection and transferring of data, and large amendments to the analysis routines due to the differences of file structures from this CCD versus the CMOS sensor.

6.9.2 Pixel Displacement Maps

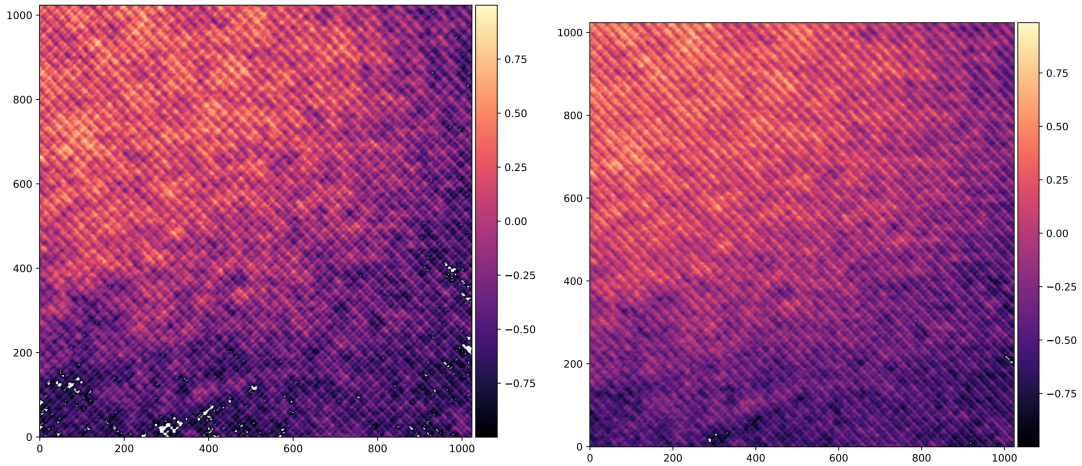
The first set of optical fringe data (45° with the fibers spaced at 2.5 mm) were fed through the analysis pipeline and pixel displacement maps obtained. Figure 6.30 shows a crop of a single frame of fringes, and the resulting spatial fit of that area. Here it clear that even a small region of fringe data shows great non-uniformity across the image, an example of the importance of flat-fielding. Figure 6.31 shows the displacement maps cropped to a cropped to a 1024×1024 pixel region to better show the smaller-scaled structure. From these results there is a clear diagonal structure to the data that resembles the original optical fringes, something not observed in the simulations of Chapter 5. This could be an indication that the spatial fit has either under or over fitted the fringe periodicity and this difference has propagated through as pixel displacements. The spread of the results are shown as a bivariate distribution in Figure 6.32. Here the displacements are positively correlated whereas the expected distribution should be a Gaussian centred about 0. This again could indicate that my spatial fitting is injecting some bias or systematic and that these measurements are not a true reflection of the detector characteristics. Also, the scale of the displacements is large, often a significant fraction of a pixel. This may have been expected for the CMOS sensor but this CCD is expected to be of higher quality fabrication with pixel displacements on the scale of milli pixels.

There are other interesting structural features in the spatial maps of Figures 6.31a and 6.31b in that the measured displacements are generally positive in the upper left of the detector, and negative in the other three corners. This, again, could be a result of an error in the analysis procedure or that the actual detector is physically warped. During the mounting procedure, the CCD bracket is pressed on all four corners to ensure a good electrical and physical connection to the socket, and that it lies perpendicular to the optical axis. As discussed in Section 6.8.1, this alignment was checked. However, some small degree of warping may be present which is propagating through the analysis as a displacement. With my current experimental set up it was difficult to further investigate this, but is noted for future experiments regarding the HARPS3 science-grade CCD.



(a) Fringe data (b) Fit to fringe data

Figure 6.30: a) The optical fringes cropped to a 256×256 pixel region, b) the image generated from the parameters of the spatial fit to that image, cropped to the same size for comparison.



(a) X Displacements (b) Y Displacements

Figure 6.31: a) The X displacements for the CCD sensor cropped to a 1024×1024 pixel region, b) the Y displacements. Any displacement greater than 1 pixel is coloured white.

These initial results are cause for concern. Despite the efforts for a robust and reliable data reduction pipeline, and an experiment designed from the ground up to be thermally stable, the pixel displacement measurements are two to three orders of magnitude greater than anticipated. The CCD was expected to have pixel displacement values of around 0.001 pixels, and the pipeline has returned some clearly non-physical solutions of nearly one pixel. There are a number of possible explanations that are expanded on in Section 6.11. But one testable avenue

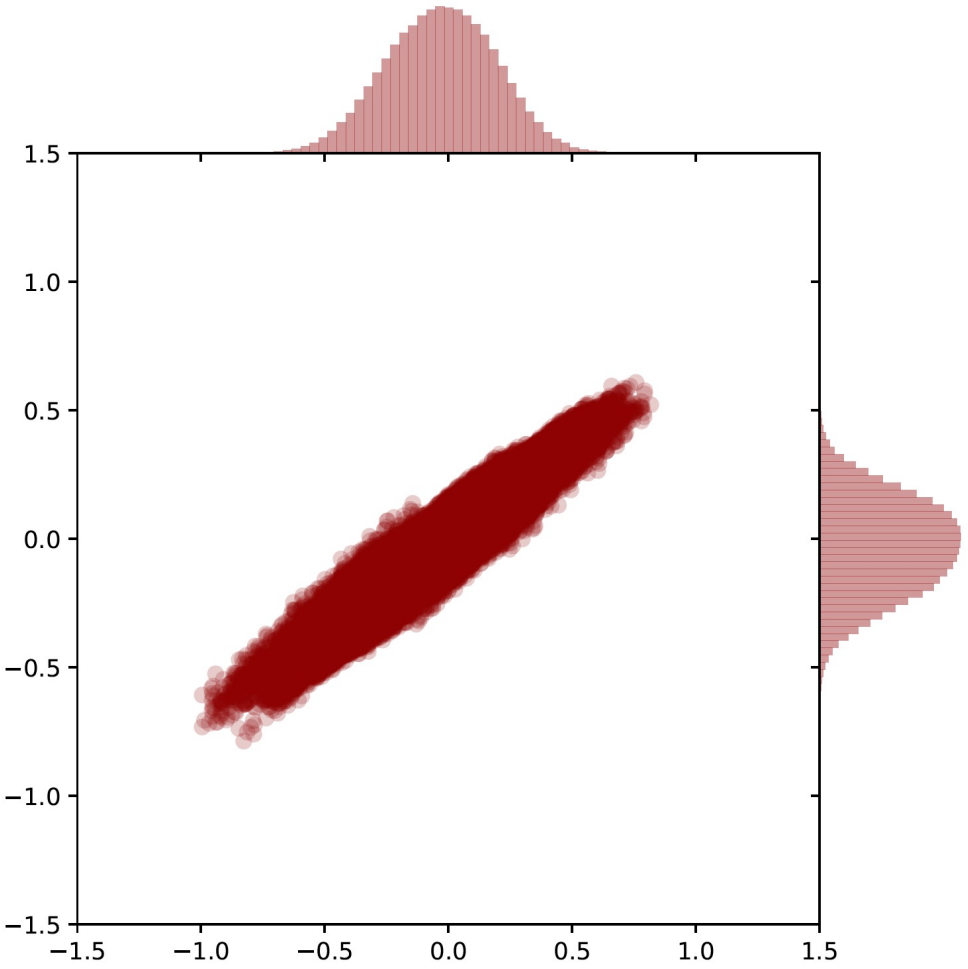


Figure 6.32: The bivariate distribution of the pixel displacements from the HARPS3 Test CCD, units are in pixels.

was to eliminate the wild temperature swings induced by the clean-room's air handling system.

6.10 Further Environmental Mitigation

One last test of whether the temperature swings can account for at least *some* of the large pixel displacement values was with a complete shutdown of the air handling system of the clean-room. The motivation for this is that as the clean-room is fairly isolated inside the greater Cavendish Laboratory, any sharp change of external temperature would take some time to materialise inside the laboratory. Hence, if the experiment could be conducted in a period of relatively low temperature swings (e.g. the middle of the day or the middle of the night) then perhaps the overall increase in stability of the laboratory will yield better results even if the average temperature is much higher. One potential downside to this is the potential increase in

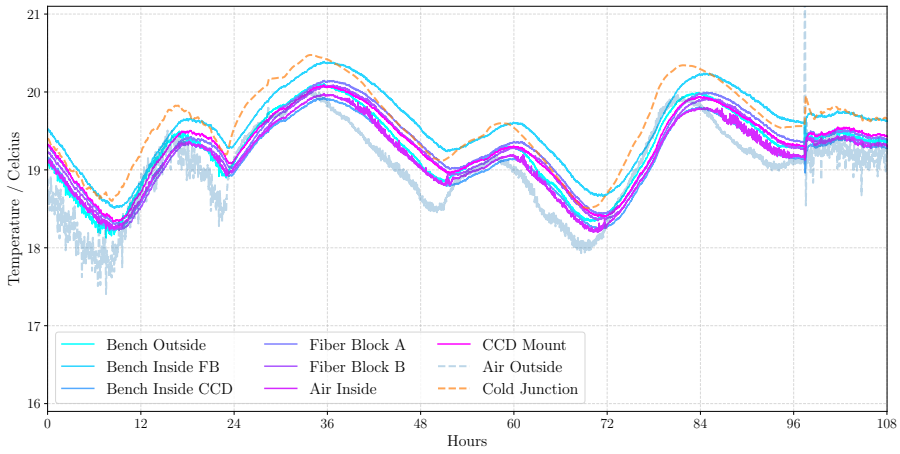


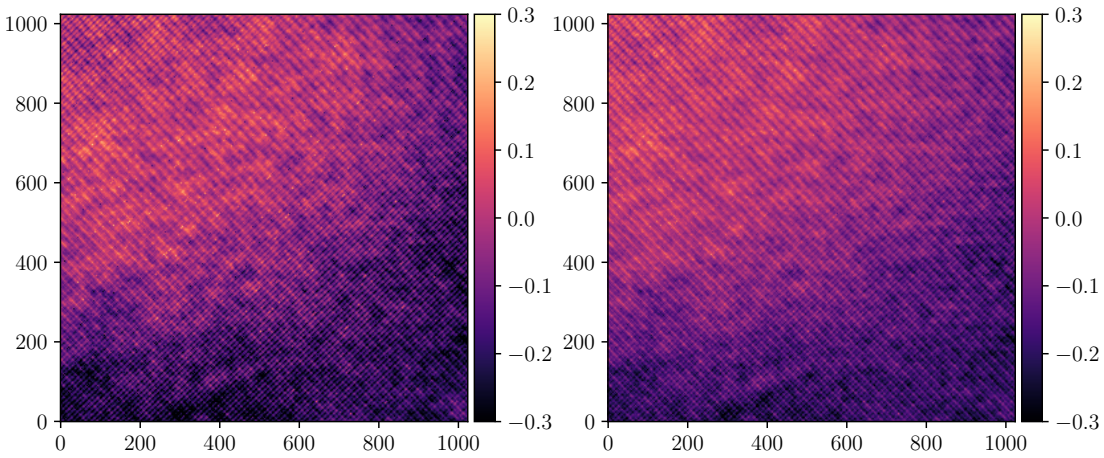
Figure 6.33: The measured temperatures at various points in and out of the blackout enclosure where the air handling is switched off for almost 72 hours over one weekend. The system is switched off at the 24-hour mark, and the large spike at 96 hours is the air handling system rebooting.

humidity, or the lack of control over it, which could have unforeseen consequences.

It was arranged for two consecutive weekends to completely shutdown the system on a Friday afternoon, and to conduct the experiment during the midday hours of both the weekend days, and then reboot the system early on a Monday such that normal business can resume for other members of nearby labs that may be affected.

Figure 6.33 shows the drastic change in short-term temperature stability that can be obtained simply by switching off the air handling system. The short period 20-minute cycles are removed, but the longer variations are amplified, likely an artefact of the outside temperature variations. During this window, a further fringe data-set was collected and processed through the pipeline to assess if this last step of environmental mitigation decreases the spread of measurements suspected to arise from an unstable experiment.

The results from the repeated experiment are shown in Figures 6.34 through 6.35. The pixel displacement maps have been cropped to the central 1024×1024 region, and the bivariate distribution of all pixel measurements is shown. It is clear here that eliminating the short-period temperature swings has reduced the spread of measurements by roughly a factor of 4. The standard deviation of the measurements in Figures 6.31a and 6.31b is 0.443 pixels and 0.368 pixels respectively, whilst in Figures 6.34a and 6.34b it is 0.104 and 0.086 pixels respectively. Note, this is not the measurement precision of the pixel positions, just the spread of the data. This is a useful metric when dealing with likely noise-dominated results, as noise mitigation is



(a) X Displacements

(b) Y Displacements

Figure 6.34: a) The X displacements for the CCD sensor cropped to a 1024×1024 pixel region, b) the Y displacements. Here the HARPS3 cleanroom temperature control is switched off to eliminate the short-period temperature swings.

clearly effective here. This reduction provides very good evidence that the measurements taken with the short-period temperature oscillations were dominated with temperature dependent noise which manifested in the results as large pixel displacements. I anticipate that further control of the environment allowing for longer and multiple data collection runs will allow for a more trustworthy measurement of the pixel positions to be made. As it stands it is impossible to say to what precision these values are, as they are a single measurement per pixel and are likely still dominated by uncharacterised noise sources.

A possibly promising result is the overall landscape of the results seem to be consistent in the datasets. Both before and after the temperature control shutdown, the large pixel maps exhibit larger positive displacements in the upper left of the frame, and larger negative displacements in the lower right of the frame. This could be a reflection of the true underlying CCD pixel structure, or evidence of a warping of the CCD due to the imperfect mount on the PCB board. More measurements of pixel displacement maps would be needed to confirm this.

To visualise the global patterns shown in Figures 6.34a and 6.34b, I have combined the datasets and plotted the 2D vector of the pixel displacement as an arrow. I have also average-binned the original 2048×2048 array into an image of size 64×64 'pixels'. Each arrow is the average vector of a group of 32×32 pixels. Here, the vectors show an overall 'rotation' of the CCD measurements slightly counter-clockwise when hinged in the top left corner. This could be indication that the CCD was not mounted securely through the data collection period, or that thermal fluctuations have caused a drift to propagate through the analysis pipeline.

Figures 6.34a to 6.36 revealed large-scale, global variations of displacements across the

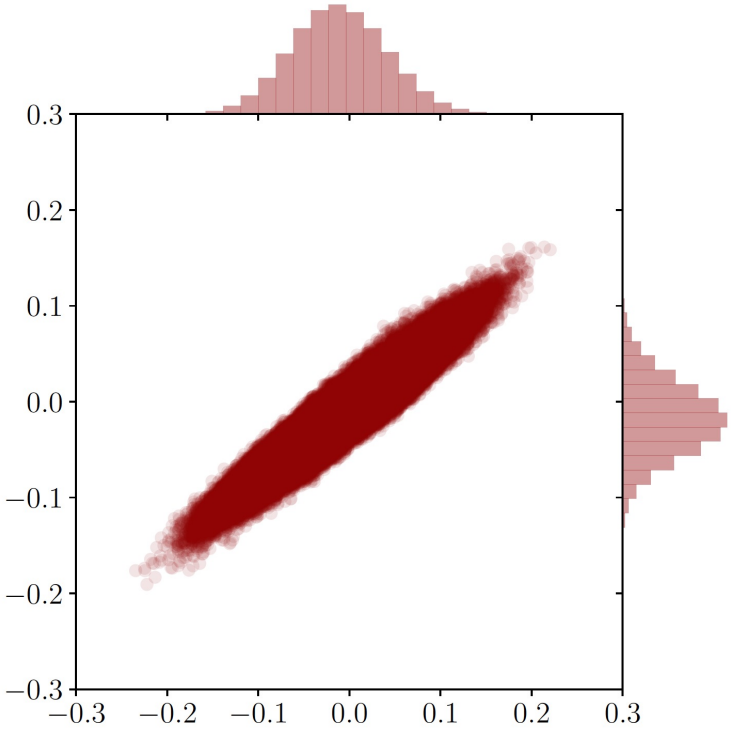
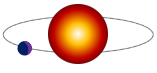


Figure 6.35: The bivariate distribution of the pixel displacements from the HARPS3 Test CCD with the HARPS3 cleanroom temperature control turned off, units are in pixels.

detector. These were likely the result of a non-perfect interference model propagating through the analysis script combined with the temperature variations witnessed in the environment. To better show the local pixel displacements, I have binned and averaged the displacements in 16×16 pixel bins, and subtracted the displacements from the value of the binned-average pixel they sit in.

This essentially acts as an analogy of a flat-field correction, but for displacement values rather than intensity. Figures 6.37 and 6.38 show the results of subtracting the average of a pixel's nearest neighbours displacement. Here, the global scale of displacements is much reduced and is a better metric for pixel to pixel displacement values. In this regime, the standard deviation is reduced to 0.053 and 0.039 pixels for the X and Y displacements respectively. These values are significantly lower than for the original displacements standard deviation of 0.105 and 0.087 pixels for the X and Y displacements respectively.



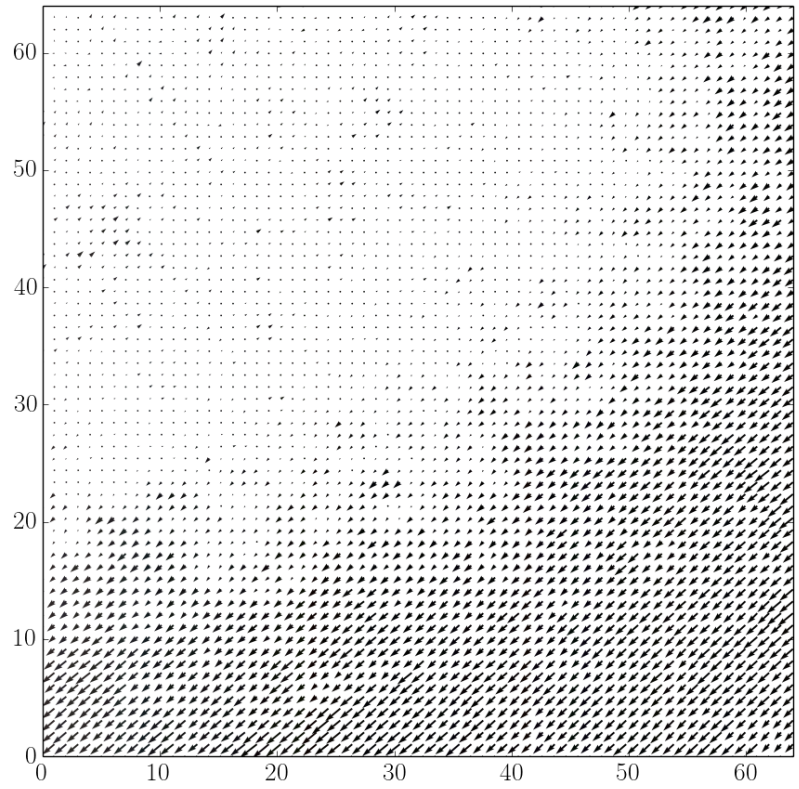


Figure 6.36: A 'quiver' plot showing the 2D vector of each pixel displacement binned into an image of 64×64 to show global displacement patterns.

6.11 Immediate Next Steps

The experiment to measure the effective pixel positions of the HARPS3 CCD to within 1 milli-pixel precision has not been successful. However, there are a number of possible reasons for this and any further experimentation along this route should consider, in no particular order, the following points.

The first, and most obvious, issues to fix would be control of the environment. The simulation and the analysis pipeline assume that the conditions on the detector are *identical* between frames. As a typical data collection window lasts a few hours, it is critical that air and optical component temperatures, lab humidity, and clean-room standards be maintained. The clean-room was not built to the specification of a ± 1 K variation within 24 hours and instead the temperatures and humidity swung wildly. It is possible that if we had the desired stability, the experiment would have provided better results. The knock-on affect of this was measured

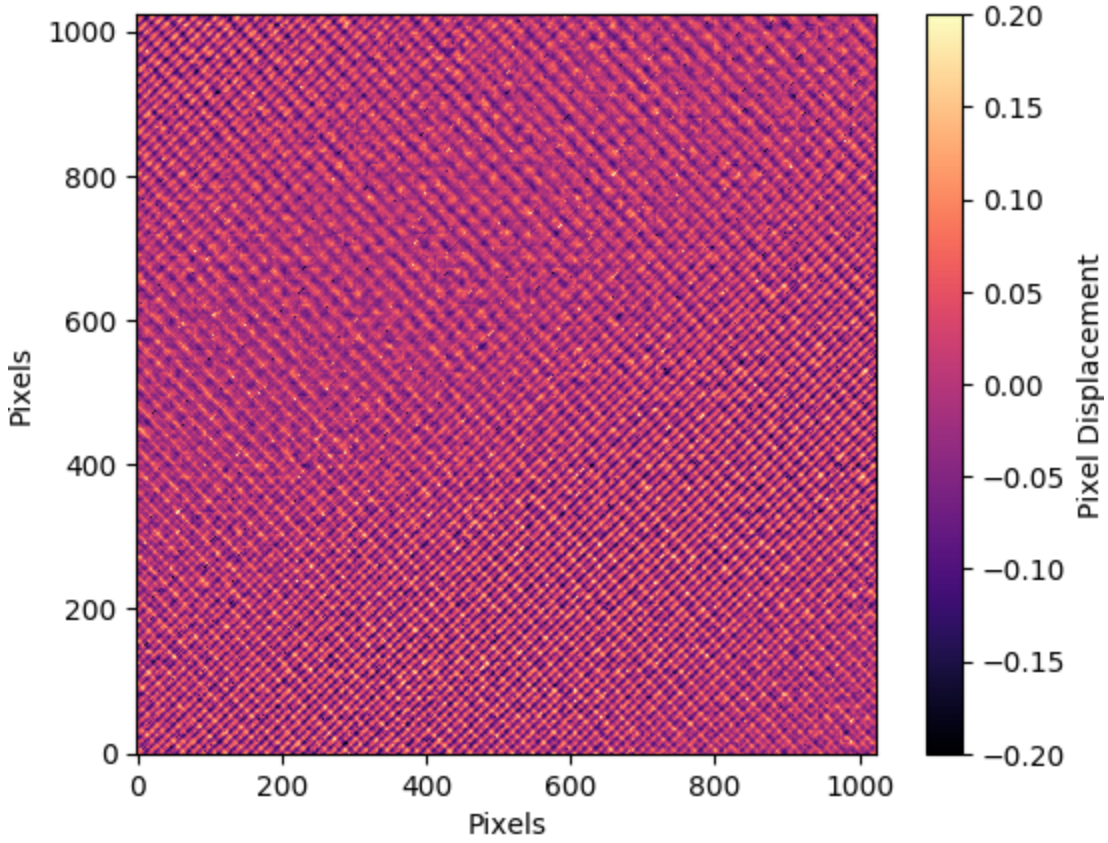
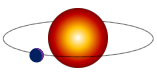


Figure 6.37: The X-displacements after averaging for global variations on a coarse 16×16 pixel grid swept across the image. Each pixel's displacement has this average bin value subtracted from it to normalise for large-scale variations.

directly on the various optical components on the optics bench. As the coefficient of thermal expansion (CTE) of the various pieces of optical equipment was different, the optical baselines were effectively 'breathing' every 20 minutes, and 'sighing' every few hours. Any relative motion between the CCD and the lasers will propagate through the analysis pipeline and the algorithm will try to deproject that as a pixel displacement.

To mitigate this further, a more stable environment could be created inside the HARPS3 clean-room constructed with heavy plastic curtains that hang from the ceiling to the floor and completely surround the optics bench. These are known to dampen the flow of air and hence the heat transfer to the enclosed section of the lab. Additionally, the baseline between the fibers and the CCD could be constructed out of a super-low CTE material such as a carbon-fibre breadboard or a ceramic such as Zerodur. Enquires into such a purchase were made, but the lead times exceeded the reach of the project.

Another limiting factor was in the choice of test detectors. The first detector used was a CMOS sensor whose results are difficult to interpret due to the inherently different pixel



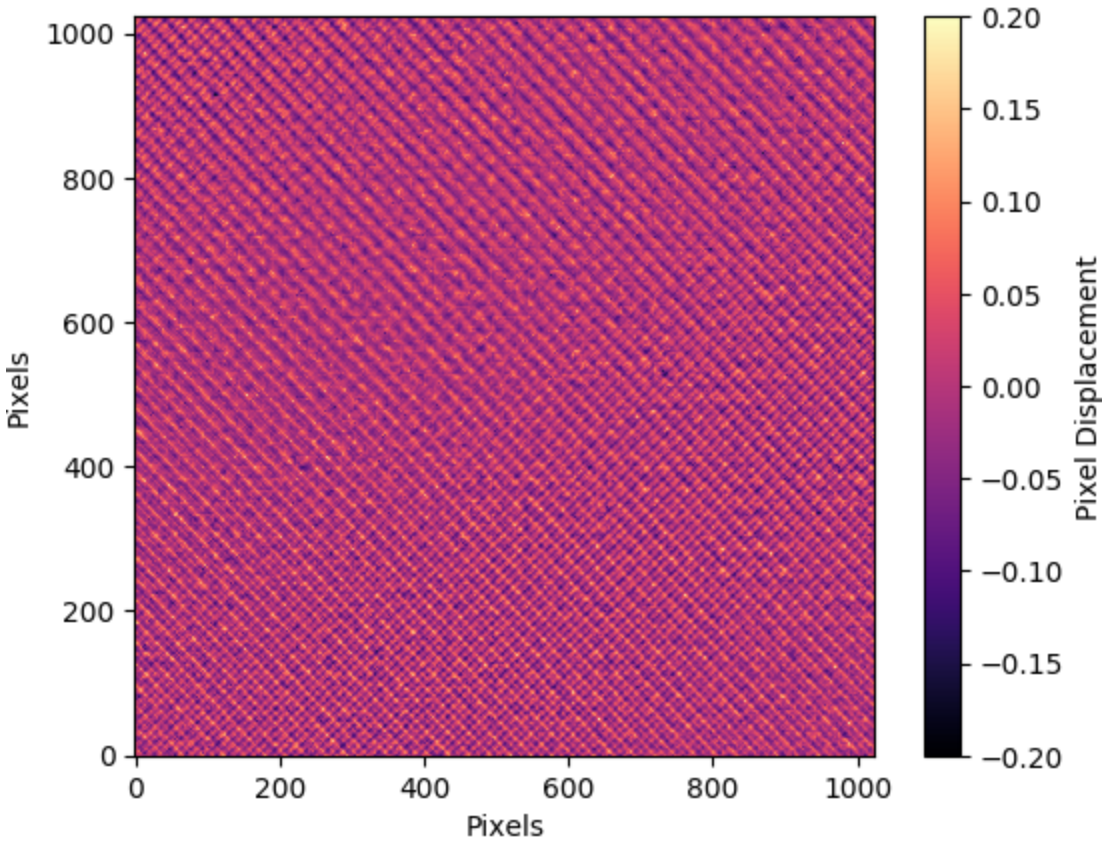


Figure 6.38: The Y-displacements after averaging for global variations on a coarse 16×16 pixel grid swept across the image. Each pixel's displacement has this average bin value subtracted from it to normalise for large-scale variations.

architecture from a CCD. The HARPS3 Test CCD (2k × 2k pixels) was of engineering grade quality. This means the detector is electronically functional but has virtually no guarantee of imaging quality. This detector was notoriously difficult to get operational: many months were spent on installation, setup, and calibration, and only in the very final stages of the experiment were we able to extract useful science frames for fringe analysis. Because this detector runs at room temperature, a lot of the standard configuration files (e.g. from HARPSN) had to be manually edited one byte at a time in a near trial-and-error fashion until it was operational. Almost all of the images collected exhibited sharp gradients in counts, some degree of blurring or smearing, and many dead/hot pixels. If the fringes collected are not a near perfect representation of the true interference pattern, these issues will also propagate through the pipeline and the code will try to extract pixel displacements regardless. Combining this with the temperature variations essentially results in near meaningless data.

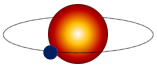
A possible next step for this would be to purchase a trusted CCD sensor of modest quality, size, and value, e.g. 1024×1024 pixels, that is pre-calibrated and ready to take science frames.

An example would be something that a keen amateur astronomer would purchase for long exposures. With a trustworthy detector, the thermal issues of the lab and how that affects the fringe deprojection could be better investigated and mitigated. Once this step is complete, *then* could the HARPS3 Test CCD be used and only after that would the HARPS3 Science CCD be measured. As it stands, the experiment has too many degrees of freedom of unknowns and uncalibrated issues.

An entire data collection run of a set baseline with the block rotated and many hundreds of frames collected will just result in one measurement of a displacement vector per pixel. Then, many tens of sets would be collected of varying baselines and fiber angles such that an estimate of both pixel position and measurement precision can be made. For a set of many measurements the spread will indicate the accuracy, precision, and repeatability of the overall experiment. Once the per-pixel position is known to some sub-milli-pixel precision due to many repeated measurements, the information can then be used to update the data reduction pipeline of the HARPS3 analysis platform and provide more accurate positions and motions of spectral lines on the detector.

When in use, the HARPS3 CCD will sit behind an optical window that forms part of the vacuum seal to the continuous flow cryostat so that the CCD can be operated in a cold and stable environment. Part of this experiment that was never completed was to conduct the pixel mapping procedure with an optical flat window placed in front of the CCD. The parameters and coefficients that define the surface profile of the optical window can be added to alter the expected interference pattern as seen by the detector. Voelz (2011) details the procedure for adding custom optics to the optical path, such as lenses and other refracting elements, which would be used in a test environment on the optical bench with the HARPS3 test CCD. Then, if the pixel displacement measurements can be shown to be repeatable with and without the optical window, this procedure will be then be used on the HARPS3 CCD in situ behind the vacuum seal window.

The last major step to a more successful experiment would be more time. Due to the fact that the HARPS3 cleanroom construction was over 18 months delayed from the initial completion date, the experiment was essentially 'on-hold' between Sections 6.1 and Section 6.2, it was during this time that the work of Chapter 3 was conducted. Unfortunately, the experiment was only operational during the final months of the allocated time and would benefit greatly from an additional 12 - 18 months of work to address the previously mentioned points and more.



Chapter 7

Overall Conclusions and the Next Steps

> 'But in the end, it is only a passing thing, this shadow. Even darkness must pass. A new day will come.'
>
> *Samwise Gamgee*

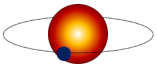
In this concluding chapter I present a summary of the three major projects undertaken during my PhD and discuss the results. I also present my ideas for how to expand on each, and discuss the imminent next steps should further work on these projects be conducted.

7.1 The Terra Hunting Experiment Feasibility Study

The aim of this project was to create a series of radial velocity datasets that are characteristic of those likely to be achieved with the Terra Hunting Experiment and others for comparison. Then, a novel nested sampling algorithm was used to fit planetary models to the data in order to compare the relative successes and failures of the various schedules.

With this framework I have demonstrated that an intense series of observations of a solar-like star can reveal the presence of an Earth-twin, and that this type of schedule outperforms the current best efforts on similar instruments. Pushing the schedule to a continuous baseline only marginally improves the detection capabilities but these schedules do improve on parameter estimation. In all cases, an idealised stellar signal mitigation was conducted.

I used SOAP2.0 to generate RVs from physical processes occurring on the surface of the star which are a common source of false positives. However, SOAP2.0 only considers the rotation



and evolution of sunspots and faculae across the stellar disk. Pressure waves, granulation, and long-term magnetic activity cycles are not considered here. An extension to SOAP could be feasibly created on the timescale of around six to nine months which would incorporate magnetic cycles by modulating the overall activity level and spot numbers.

Also, for all of the solar systems tested, the exact same SOAP data was used for practicality and comparison, and the planets in each were manually generated and not tested for dynamic instability. This code-base could be used as a foundation for a full Terra Hunting Yield estimator that incorporates a variety of stellar RV data sets, and a verified stable solar system that may or may not contain an Earth-twin. From here, the scheduling of observations could be updated to consider a full 10-year set of unique time-stamps on a variety of targets for a full survey. This would enable the true yield of a schedule to be tested, instead of only testing the feasibility of an idealised schedule.

Lastly, as the current implementation of end-to-end code is self-contained it would be feasible to package it as a public tool for the exoplanet community. A user would only need to define a few parameters such as observation schedule time-stamps, planetary systems to include, and SOAP2.0 parameters. Then the code would execute the nested sampling with PolyChord and return the posterior distributions and model evidences. A user could trial a variety of proposed schedules and SOAP2.0 RV series to investigate how their proposed schedule is hindered by different stellar RV series and tune their schedule for optimum results. However, creating this package would be a lengthy project that would likely take around 12 months and hence was not part of the scope of the research.

7.2 The Pixel Mapping Simulation

For this project I created an end-to-end simulation of an optical detector that is imaging optical fringes from two small sources of monochromatic light originating from optical fibers. The detector is given small per-pixel displacements and the resulting intensity is sampled. By analysing the measured signal, the injected pixel displacements can be measured under the assumption the fringes are a constant ‘optical ruler’. This project served as a precursor to the experiment being simulated that would eventually be used to measure the pixel positions of the HARPS3 detector, and helped to guide the experimental construction and design.

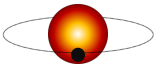
With this simulation I investigated a wide range of variables such as the size of the detector, the level of photon noise, pixel stitching boundaries, the number of measurements, fiber baseline spacing, and fiber baseline angle. This broad set of parameters allows an experiment to be designed to achieve optimal results and can be used to simulate the pixel mapping experiment on practically any optical detector.

The simulation does make a few assumptions however which could be rectified with further work. Firstly the injected pixel displacements are drawn from a uniform distribution which may not be physically realistic. Alternate distributions, such as a Beta distribution, may better describe the true displacements that arise from the fabrication of the detector. Regardless, a next step would be to verify that the analysis pipeline can handle different types of initial displacement distributions which could mean using alternate curve-fitting routines than just a standard least-squares minimisation.

Also, recent work has revealed that optical detectors also exhibit intrapixel variations in photometric sensitivity, i.e. the pixel will register different counts depending on *where* the photons strike within a pixel's boundaries. Ravi et al. (2017) used a laser frequency comb to aid in the characterisation of the HARPSN spectrograph and achieved a $1\,\mathrm{cm\,s^{-1}}$ calibration source stability. They also reported that they can measure the intensity variations of the comb to much less than one pixel. These small variations are assumed to be zero in my simulation and data analysis pipeline. It is feasible that the interference fringes can be used to probe the intrapixel intensity variations. But this would require a sizeable branch of the simulation to investigate. An LFC is an ideal tool for this as it comprises of a series of regularly spaced intensity peaks, a feature shared with the optical fringes, so it is likely that intra-pixel variations could be investigated here. The simulation also assumes that dark and flat-field corrections have already been applied such that the data is assumed to only contain optical fringe information.

The presence of lithographic blocks was largely ignored in the analysis but with a priori knowledge of the block size, the block displacement vector could be incorporated into the analysis to allow the fitting routine to better find a solution. This would be a fairly involved amendment to the pipeline as the simulation and generation of optical fringes from the basis of curve fitting are fundamentally relying on a set of defined points used to sample the interference pattern.

Lastly, the HARPS3 detector (and likely any others used in a low temperature spectrograph) will sit behind an optical window that forms part of the seal of the vacuum chamber of the cryostat. However, any light passing through this optical window will be subject to small distortions. The procedure used to simulate the optical fringes, Voelz (2011), goes on to describe a technique of modelling optical elements such as mirrors, lenses, and refracting windows as a multi-term polynomial. Should such a window be required, the coefficients of this polynomial are usually supplied with the specification datasheet of the manufacturer and hence the effect of the window on the optical fringes can be incorporated into the analysis pipeline.



7.3 The Pixel Mapping Experiment

The experiment generated optical fringes as predicted by the optical simulation. Both the estimation of optical losses and the structure of the fringes (periodicity, intensity, uniformity) agreed well. This indicates that the formulations and approximations set out by Voelz (2011) are appropriate for this use case.

Unfortunately the experiment was hampered by uncontrollable temperature variations that plagued the stability of the data collection. The analysis assumes the fringe angle and periodicity are static throughout a single data collection period. Any changes to this will result in poor modelling of the underlying structure and these errors propagate as large pixel displacements. The analysis also assumes that the CCD and the fibers are in a fixed reference frame and do not move relative to each other. Should the optic bench gently expand and contract due to the small temperature variations witnessed then the pixel displacement measurements are redundant.

I did not have the facilities of Crouzier (2014), such as mK stability and low thermal expansion materials, with which to conduct the experiment. And the measures I took to mitigate the temperature variations were clearly not enough. However, I still believe the experiment and analysis pipeline *does* have the capability to measure pixel positions to less than 1 millipixel and could yield results comparable to Shaklan et al. (1995) and Crouzier (2014), see Section 5.3 onwards. This is due to a more sophisticated optical model and analysis pipeline, more optical baselines available, and a completely in-fiber optical experiment.

The immediate next steps to verify this claim would be to further eliminate the susceptibility of temperature variations. The fiber baseline block could be manufactured from a low thermal expansion material such as Zerodur or carbon-fiber as these exhibit expansions many times lower per degree than aluminium. Construction the 1.5 m baseline from Zerodur would be too expensive and difficult to manufacture, but the aluminium surface could be heated to a temperature above the variations witnessed in a lab and held there with a feedback loop, e.g. 25°C. The optical bench could have a number of heating elements and calibrated theromocouples mounted at various places, and the feedback loop would act to stabilise the entire surface of the bench at the set point. It is likely that these two steps alone would result in an increase in pixel mapping precision. The optical fibers are also subject to the temperature variations but the added temperature stability of the optical table would likely mitigate these as they are physically mounted to the optical table.

Should the experiment be at a stage where it is generating repeatable pixel displacements, there are a few extra steps that would be a satisfactory expansion to this work. As mentioned in Section 7.2, the presence of an optical window will cause some distortion to the fringes even if it is of high quality flatness. This is because the fringes propagate radially from the source

whilst the window is flat and sits perpendicularly to the propagation axis. Fringes towards the edge of the CCD will exhibit more refraction than those at the centre which results in a non-uniform distortion. Plus, the window itself will not be perfectly flat. It will have small deviations of around $\lambda/10$ across its surface and may have additional curvature, warping, and bending. Factoring in how the fringes are affected with the window to be used on the HARPS3 CCD could be critical for in-situ pixel displacement measurements. The window would need to be measured and analysed, and the surface properties factored in to the data analysis.

The experiment was also designed to be optically efficient for all wavelengths in the HARPS3 spectral range (380 nm to 690 nm), and the pixel structure may not be uniform at each wavelength. Photons of different wavelengths will penetrate the silicon at different depths, and the insulating walls of the pixel are likely to deviate on some small scale. Measuring the displacements at multiple wavelengths in this bandwidth would provide a more accurate wavelength solution for the HARPS3 data reduction pipeline.

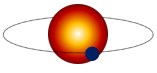
Lastly, the experiment itself could be made portable. The section containing the laser source to the optical block holding the fiber baselines could easily be installed in a miniature stabilised enclosure that acts as a 'black box' with a few USB cables protruding. This would enable the set up to be brought to existing spectrographs, such as HARPS, HARPSN, or ESPRESSO, and conduct the pixel mapping procedure on site. If fully automated, the data collection for all available baselines would only take a small number of days and could be conducted in the daytime for minimal disturbance to the night time operations of the instrument. Should the technique be proven to be accurate and reliable, this service may be of interest to existing instruments around the world.

7.4 Summary

In summary, I have conducted three major projects to aid in the discovery of a so-far elusive Earth-twin exoplanet.

I have created a new technique of analysing synthetic radial velocity data that provides full Bayesian model comparison and parameter estimation in a computationally viable way. This framework is currently being expanded to include the analysis of the spectral indicators used to infer the presence of stellar variability which proved to be a limiting factor in the simulations ability to detect the lowest-mass planets.

I then developed a new technique to measure the effective pixel positions of an optical detector to be used in a highly accurate spectrograph, HARPS3. This technique involved creating an end-to-end optical simulation of an experiment and writing the data analysis scripts from the ground up. The simulation then inferred the optimum design of an optical experiment which I



constructed and tested on a 2048×2048 pixel CCD that resembles the HARPS3 science CCD. Whilst environmental effects ultimately hindered the precise measurements, the experiment and analysis procedure will have a promising future should some minor improvements be made.

This work is just a small but important piece of the puzzle that is currently being solved by astronomers, engineers, and data scientists all over the world. We all share the common goal of the discovery of an Earth-twin and it will be a monumental success for the exoplanet community if it is made. Whilst we are confident in a detection in the near future, our efforts may instead show that Earth like planets are extremely rare, then we will learn how precious our planet truly is. Arthur C. Clarke summarised these outcomes succinctly with 'Two possibilities exist: either we are alone in the Universe or we are not. Both are equally terrifying.'.

FEASIBILITY STUDY RESULTS

In this appendix I present the tabulated results from the Terra Hunting Experiment Feasibility study. For convenience, I have denoted the schedules either REF, THE, or SPACE to represent the reference, Terra Hunting, and Space schedules respectively. I then append the duration after such that, for example, THE_10 corresponds to the 10-year Terra Hunting schedule.

Tables A.1 to A.4 displays the results where the SOAP data has been omitted and PolyChord is free to explore all planetary models.

Table A.5 displays the results where the SOAP data has been included, but PolyChord has been constrained to the correct number of planets.

Tables A.6 to A.9 displays the results where the SOAP data has been included, and PolyChord has been unconstrained to search all planetary models.

Table A.1: Model evidences and parameter estimates for System 1. PolyChord is not constrained in the choice of model, only the parameters of only the favoured model are shown. For clarity the favoured model has been underlined. The data only contains planetary RVs and Gaussian noise. $|\Delta \ln R|$ has the zero–point set to the favoured model. The correct model is indicated with a check mark. The results are discussed in section 3.7.1.

	$\Delta \ln R_{\text{favoured}} = 0$							
Schedule	$N_p = 0$	$N_p = 1$	$N_p = 2$	$N_p = 3$	$N_p = 4$	RV Semi Amplitude / $\mathrm{m\,s^{-1}}$	Period / days	Phase / radians
True		✓				0.11	293	3.46
REF_5	-4.48	<u>0.00 ± 0.25</u>	-3.14 ± 0.20	-6.16 ± 0.19	-9.56 ± 0.18	0.13 ± 0.03	292.70 ± 82.17	3.23 ± 0.47
REF_10	-9.00	<u>0.00 ± 0.28</u>	-3.93 ± 0.23	-6.70 ± 0.20	-11.13 ± 0.21	0.11 ± 0.02	272.24 ± 70.07	3.08 ± 0.78
THE_5	-18.68	<u>0.00 ± 0.30</u>	-3.32 ± 0.24	-6.27 ± 0.21	-10.96 ± 0.23	0.11 ± 0.01	286.47 ± 7.45	2.88 ± 0.28
THE_10	-37.24	<u>0.00 ± 0.30</u>	-5.53 ± 0.24	-8.76 ± 0.21	-14.83 ± 0.23	0.10 ± 0.01	290.89 ± 1.25	3.13 ± 0.19
Space_5	-51.24	<u>0.00 ± 0.27</u>	-5.68 ± 0.26	-9.30 ± 0.23	-13.41 ± 0.22	0.11 ± 0.01	290.17 ± 2.16	3.28 ± 0.17
Space_10	-109.84	<u>0.00 ± 0.29</u>	-8.52 ± 0.29	-11.42 ± 0.29	-19.32 ± 0.26	0.11 ± 0.01	292.10 ± 0.08	3.41 ± 0.12

Table A.2: As per Table A.1 for System 2, the results are discussed in section 3.7.2.

	$\Delta \ln R_{\text{favoured}} = 0$							
Schedule	$N_p = 0$	$N_p = 1$	$N_p = 2$	$N_p = 3$	$N_p = 4$	RV Semi Amplitude / $ms^{-}1$	Period / days	Phase / radians
True				✓		0.11	197	3.46
						0.11	293	4.45
						10.34	2953	1.83
REF_5	-89,216.94	-1.61 ± 0.37	0.00 ± 0.28	-0.58 ± 0.28	-1.11 ± 0.28	0.27 ± 0.50	864.46 ± 929.07	3.39 ± 1.43
						10.33 ± 0.18	2947.92 ± 25.02	1.81 ± 0.04
REF_10	-157,854.40	-9.10 ± 0.38	-2.93 ± 0.34	-0.18 ± 0.38	0.00 ± 0.32	0.07 ± 0.02	159.37 ± 56.51	3.11 ± 0.92
						0.09 ± 0.02	341.13 ± 378.18	3.66 ± 1.12
						0.97 ± 2.92	1543.42 ± 565.01	1.33 ± 0.90
						9.50 ± 2.82	2990.96 ± 138.74	2.14 ± 1.06
THE_5	-266,916.05	-41.56 ± 0.38	-11.88 ± 0.38	0.00 ± 0.47	-3.82 ± 0.46	0.12 ± 0.04	286.03 ± 117.44	2.89 ± 1.38
						0.20 ± 0.38	767.69 ± 686.73	3.64 ± 1.68
						10.30 ± 0.31	2963.98 ± 36.68	1.84 ± 0.13
THE_10	-476,030.22	-83.23 ± 0.40	-31.29 ± 0.36	0.00 ± 0.73	-1.49 ± 0.94	0.10 ± 0.01	197.45 ± 1.18	3.58 ± 0.20
						0.12 ± 0.01	294.78 ± 1.96	4.77 ± 0.20
						10.35 ± 0.01	2954.01 ± 1.44	1.83 ± 0.01
Space_5	-572,825.73	-102.19 ± 0.38	-45.38 ± 0.39	0.00 ± 0.42	-2.08 ± 0.52	0.10 ± 0.01	198.57 ± 1.14	3.62 ± 0.19
						0.11 ± 0.01	292.82 ± 2.33	4.51 ± 0.17
						10.35 ± 0.01	2957.53 ± 3.12	1.83 ± 0.01
Space_10	-1,026,898.77	-200.65 ± 0.40	-89.99 ± 0.42	0.00 ± 0.48	-1.49 ± 0.86	0.10 ± 0.01	196.66 ± 0.41	3.37 ± 0.14
						0.11 ± 0.01	293.27 ± 0.81	4.58 ± 0.13
						10.34 ± 0.01	2955.01 ± 0.84	1.83 ± 0.01

Table A.3: As per Table A.1 for System 3, the results are discussed in section 3.7.3.

	$\Delta \ln R_{\text{favoured}} = 0$							
Schedule	$N_p = 0$	$N_p = 1$	$N_p = 2$	$N_p = 3$	$N_p = 4$	RV Semi Amplitude / m s^{-1}	Period / days	Phase / radians
True				✓		0.16	101	3.46
						0.13	197	4.45
						0.11	293	1.83
REF_5	-11.75	-8.09 ± 0.27	-1.57 ± 0.25	-0.75 ± 0.33	0.00 ± 0.30	0.16 ± 0.04	99.08 ± 9.87	3.30 ± 0.46
						0.10 ± 0.03	167.30 ± 38.65	3.83 ± 1.67
						0.12 ± 0.03	248.01 ± 51.82	3.52 ± 1.76
						0.11 ± 0.03	508.95 ± 231.93	3.33 ± 1.53
REF_10	-27.01	-18.74 ± 0.27	-6.07 ± 0.45	-1.91 ± 0.35	0.00 ± 0.33	0.16 ± 0.02	74.80 ± 28.77	4.72 ± 1.14
						0.12 ± 0.02	196.61 ± 1.52	4.53 ± 0.54
						0.09 ± 0.02	272.01 ± 36.15	2.30 ± 1.15
						0.07 ± 0.02	526.09 ± 195.16	2.91 ± 1.48
THE_5	-55.79	-42.65 ± 0.29	-17.69 ± 0.29	0.00 ± 0.36	-1.21 ± 0.27	0.15 ± 0.01	100.83 ± 0.29	3.46 ± 0.19
						0.12 ± 0.02	196.18 ± 1.48	4.40 ± 0.25
						0.11 ± 0.02	294.97 ± 3.76	1.85 ± 0.27
THE_10	-139.02	-104.55 ± 0.30	-44.00 ± 0.35	0.00 ± 0.36	-9.70 ± 0.97	0.15 ± 0.01	101.14 ± 0.10	3.59 ± 0.14
						0.12 ± 0.01	196.73 ± 0.51	4.47 ± 0.17
						0.11 ± 0.01	292.32 ± 1.24	1.71 ± 0.20
Space_5	-213.33	-124.11 ± 0.31	-52.09 ± 0.46	0.00 ± 0.30	-2.14 ± 0.38	0.16 ± 0.01	100.66 ± 0.18	3.30 ± 0.12
						0.13 ± 0.01	195.86 ± 0.90	4.32 ± 0.15
						0.11 ± 0.01	292.81 ± 2.32	1.78 ± 0.17
Space_10	-444.34	-258.80 ± 0.33	-118.82 ± 0.50	0.00 ± 0.30	-18.99 ± 0.66	0.16 ± 0.01	100.98 ± 0.07	3.43 ± 0.09
						0.12 ± 0.01	196.94 ± 0.33	4.46 ± 0.11
						0.11 ± 0.01	292.46 ± 0.80	1.78 ± 0.12

Table A.4: As per Table A.1 for System 4, the results are discussed in section 3.7.4.

	$\Delta \ln R_{\text{favoured}} = 0$							
Schedule	$N_p = 0$	$N_p = 1$	$N_p = 2$	$N_p = 3$	$N_p = 4$	RV Semi Amplitude / m s^{-1}	Period / days	Phase / radians
True	✓							
REF_5	0.00	-27.43 ± 0.20	-30.89 ± 0.19	-33.64 ± 0.18	-37.18 ± 0.17	–	–	–
REF_10	0.00	-41.00 ± 0.20	-44.85 ± 0.19	-47.83 ± 0.19	-51.17 ± 0.19	–	–	–
THE_5	-3.95	0.00 ± 0.25	-3.32 ± 0.24	-8.90 ± 0.22	-13.09 ± 0.20	0.06 ± 0.01	162.80 ± 107.97	2.21 ± 1.52
THE_10	-3.27	0.00 ± 0.26	-4.35 ± 0.22	-10.45 ± 0.21	-13.90 ± 0.31	0.05 ± 0.01	240.09 ± 155.15	3.13 ± 2.42
Space_5	-35.18	0.00 ± 0.25	-4.94 ± 0.23	-10.08 ± 0.12	-15.23 ± 0.20	0.05 ± 0.01	169.66 ± 50.58	3.20 ± 0.89
Space_10	-9.89	0.00 ± 0.28	-1.11 ± 0.22	-8.59 ± 0.21	-12.54 ± 0.23	0.05 ± 0.01	52.74 ± 56.10	3.65 ± 1.10

Table A.5: Parameter estimates for all schedules observing the first three Systems where the data includes planetary RVs, Gaussian noise and quasi–periodic stellar noise. PolyChord has been constrained to the true number of planets. The results are discussed in section 3.8.

	System 1			System 2			System 3		
Schedule	RV Semi Amplitude / $ms^{-}1$	Period / days	Phase / radians	RV Semi Amplitude / $ms^{-}1$	Period / days	Phase / radians	RV Semi Amplitude / $ms^{-}1$	Period / days	Phase / radians
True	0.11	293	3.46	0.11	197	3.46	0.16	101	3.46
				0.11	293	4.45	0.13	197	4.45
				10.34	2953	1.82	0.11	293	1.82
REF_5	0.16 ± 0.03	210.13 ± 69.27	1.14 ± 0.42	0.11 ± 0.03	49.20 ± 28.09	2.03 ± 1.32	0.20 ± 0.03	76.25 ± 11.29	2.32 ± 2.77
				0.19 ± 0.11	267.74 ± 295.75	5.88 ± 0.92	0.21 ± 0.05	203.33 ± 25.85	0.64 ± 1.33
				10.38 ± 0.10	2956.10 ± 23.78	1.84 ± 0.01	0.12 ± 0.05	636.77 ± 470.34	2.46 ± 1.42
REF_10	0.12 ± 0.02	302.69 ± 65.13	3.01 ± 0.44	0.10 ± 0.02	27.50 ± 16.94	2.87 ± 0.97	0.16 ± 0.02	51.32 ± 24.55	1.20 ± 1.81
				0.10 ± 0.41	198.67 ± 175.24	1.29 ± 0.85	0.16 ± 0.02	342.78 ± 133.52	2.92 ± 0.96
				10.33 ± 0.41	2968.59 ± 30.95	1.85 ± 0.06	0.14 ± 0.02	1250.58 ± 384.32	3.95 ± 0.92
THE_5	0.13 ± 0.02	283.94 ± 3.01	2.44 ± 0.22	0.10 ± 0.02	20.88 ± 2.05	5.71 ± 0.42	0.15 ± 0.01	100.88 ± 0.35	3.25 ± 0.21
				0.11 ± 0.02	206.65 ± 9.69	4.69 ± 0.30	0.10 ± 0.02	203.30 ± 12.47	5.26 ± 0.55
				10.36 ± 0.02	2944.64 ± 4.66	1.83 ± 0.01	0.14 ± 0.02	294.94 ± 22.42	1.81 ± 0.30
THE_10	0.11 ± 0.01	290.76 ± 1.08	2.77 ± 0.17	0.07 ± 0.01	80.38 ± 33.64	4.70 ± 1.30	0.15 ± 0.01	100.88 ± 0.10	3.25 ± 0.14
				0.08 ± 0.01	291.19 ± 18.56	4.45 ± 0.43	0.10 ± 0.01	197.58 ± 7.39	4.85 ± 0.30
				10.35 ± 0.01	2959.08 ± 1.43	1.84 ± 0.01	0.14 ± 0.01	292.89 ± 10.46	1.70 ± 0.14
Space_5	0.12 ± 0.01	284.04 ± 2.12	2.67 ± 0.15	0.11 ± 0.01	202.20 ± 1.05	4.09 ± 0.19	0.16 ± 0.01	100.64 ± 0.20	3.23 ± 0.13
				0.09 ± 0.01	288.47 ± 3.78	4.01 ± 0.31	0.11 ± 0.01	200.29 ± 1.08	4.89 ± 0.16
				10.36 ± 0.01	2958.52 ± 3.08	1.83 ± 0.01	0.13 ± 0.01	291.80 ± 1.83	1.75 ± 0.13
Space_10	0.10 ± 0.01	291.73 ± 0.80	3.07 ± 0.13	0.09 ± 0.01	196.99 ± 0.48	3.31 ± 0.17	0.06 ± 0.01	23.05 ± 0.80	3.54 ± 0.23
				0.08 ± 0.01	293.26 ± 1.27	4.49 ± 0.19	0.16 ± 0.01	100.86 ± 0.18	3.34 ± 0.09
				10.34 ± 0.01	2955.60 ± 0.84	1.83 ± 0.01	0.14 ± 0.01	292.02 ± 0.77	1.67 ± 0.10

Table A.6: Model evidences and parameter estimates for System 1, per observation schedule, with the data containing Stellar Noise, Gaussian noise and planetary RVs. PolyChord has not been constrained to any model. The results are discussed in section 3.9.1.

	$\Delta \ln R_{\text{favoured}} = 0$					RV Semi Amplitude / $\mathrm{m\,s^{-1}}$	Period / days	Phase / radians
Schedule	$N_p = 0$	$N_p = 1$	$N_p = 2$	$N_p = 3$	$N_p = 4$			
True		✓				0.11	293	3.46
REF_5	-19.14	-8.75 ± 0.26	-5.87 ± 0.24	-3.05 ± 0.29	0.00 ± 0.23	0.12 ± 0.03	25.63 ± 16.65	4.55 ± 1.68
						0.12 ± 0.04	46.30 ± 51.69	1.38 ± 0.65
						0.13 ± 0.03	236.81 ± 120.35	1.64 ± 1.10
						0.11 ± 0.03	3399.24 ± 500.80	3.16 ± 0.38
REF_10	-32.59	-19.39 ± 0.29	-7.97 ± 0.25	-3.66 ± 0.27	0.00 ± 0.26	0.08 ± 0.02	77.30 ± 60.32	3.11 ± 1.76
						0.10 ± 0.04	223.14 ± 67.84	1.64 ± 0.87
						0.14 ± 0.04	425.90 ± 181.64	3.04 ± 1.53
						0.11 ± 0.03	1344.35 ± 503.66	4.38 ± 1.84
THE_5	-54.97	-25.70 ± 0.28	-15.80 ± 0.26	-8.83 ± 0.36	0.00 ± 0.44	0.08 ± 0.01	17.98 ± 2.40	2.61 ± 1.69
						0.09 ± 0.02	24.80 ± 11.11	1.37 ± 0.50
						0.13 ± 0.02	285.30 ± 6.94	2.61 ± 0.27
						0.09 ± 0.02	395.91 ± 14.34	3.12 ± 0.36
THE_10	-63.48	-18.48 ± 0.30	-7.19 ± 0.43	0.00 ± 0.48	-0.48 ± 0.35	0.07 ± 0.01	60.64 ± 19.03	1.44 ± 2.22
						0.08 ± 0.02	217.82 ± 39.22	0.98 ± 0.93
						0.10 ± 0.02	453.28 ± 343.07	3.27 ± 1.09
Space_5	-93.17	-34.77 ± 0.28	-14.56 ± 0.30	-6.27 ± 0.48	0.00 ± 0.41	0.08 ± 0.01	21.98 ± 0.02	1.30 ± 0.29
						0.06 ± 0.01	27.47 ± 9.65	1.91 ± 0.46
						0.06 ± 0.01	37.70 ± 18.49	2.91 ± 0.29
						0.12 ± 0.01	283.85 ± 8.96	2.65 ± 0.16
Space_10	-132.44	-37.66 ± 0.30	-22.15 ± 0.32	-9.98 ± 0.34	0.00 ± 0.53	0.06 ± 0.01	22.13 ± 0.33	2.84 ± 0.41
						0.05 ± 0.01	26.00 ± 0.15	6.06 ± 0.32
						0.06 ± 0.01	109.24 ± 0.25	1.51 ± 0.28
						0.10 ± 0.01	291.90 ± 0.84	3.08 ± 0.13

Table A.7: As in TableA.6 but for System 2. The results are discussed in section 3.9.2

	$\Delta \ln R_{\text{favoured}} = 0$							
Schedule	$N_p = 0$	$N_p = 1$	$N_p = 2$	$N_p = 3$	$N_p = 4$	RV Semi Amplitude / $\mathrm{m\,s^{-1}}$	Period / days	Phase / radians
True				✓		0.11	197	3.46
						0.11	293	4.45
						10.34	2953	1.83
REF_5	-89,071.19	-17.39 ± 0.36	-4.92 ± 0.33	-0.79 ± 0.37	0.00 ± 0.33	0.13 ± 0.04	44.32 ± 26.74	2.11 ± 1.36
						0.15 ± 0.04	168.51 ± 117.21	4.62 ± 1.94
						0.31 ± 1.28	1541.14 ± 402.76	1.20 ± 0.87
						10.17 ± 1.27	2995.67 ± 76.40	1.86 ± 0.08
REF_10	-157,356.78	-9.20 ± 0.38	-4.90 ± 0.42	0.00 ± 0.34	-0.92 ± 0.44	0.10 ± 0.02	27.50 ± 16.94	2.87 ± 0.97
						0.10 ± 0.41	198.67 ± 175.24	1.29 ± 0.85
						10.33 ± 0.41	2968.59 ± 30.95	1.85 ± 0.06
THE_5	-267,224.20	-38.88 ± 0.37	-20.60 ± 0.37	-7.85 ± 0.39	0.00 ± 0.88	0.10 ± 0.01	22.05 ± 0.50	1.60 ± 0.38
						0.10 ± 0.02	36.28 ± 46.11	0.96 ± 0.57
						0.10 ± 0.09	321.02 ± 42.21	5.17 ± 1.06
						10.33 ± 0.09	2953.60 ± 9.79	1.83 ± 0.03
THE_10	-475,737.51	-42.91 ± 0.39	-24.27 ± 0.33	-10.93 ± 0.38	0.00 ± 0.56	0.07 ± 0.01	55.14 ± 31.41	2.87 ± 2.22
						0.07 ± 0.01	173.14 ± 1.57	2.35 ± 0.48
						0.08 ± 0.01	295.56 ± 4.42	4.68 ± 0.53
						10.35 ± 0.01	2958.42 ± 1.97	1.84 ± 0.01
Space_5	-572,795.30	-101.15 ± 0.39	-57.53 ± 0.38	-21.51 ± 037	0.00 ± 0.38	0.08 ± 0.01	21.97 ± 0.92	1.03 ± 0.35
						0.11 ± 0.01	202.20 ± 1.09	4.09 ± 0.19
						0.09 ± 0.01	288.41 ± 3.82	4.01 ± 0.31
						10.36 ± 0.01	2958.52 ± 3.03	1.83 ± 0.01
Space_10	-1,027,062.81	-130.12 ± 0.40	-65.05 ± 0.36	-13.89 ± 0.54	0.00 ± 1.01	0.05 ± 0.01	49.35 ± 54.71	2.86 ± 0.37
						0.09 ± 0.01	196.95 ± 0.94	3.29 ± 0.18
						0.08 ± 0.01	293.23 ± 1.53	4.48 ± 0.18
						10.34 ± 0.01	2955.55 ± 0.82	1.83 ± 0.11

Table A.8: As in TableA.6 but for System 3. The results are discussed in section 3.9.3

	$\Delta \ln R_{\text{favoured}} = 0$							
Schedule	$N_p = 0$	$N_p = 1$	$N_p = 2$	$N_p = 3$	$N_p = 4$	RV Semi Amplitude / $m\,s^{-1}$	Period / days	Phase / radians
True				✓		0.16	101	3.46
						0.13	197	4.45
						0.11	293	1.83
REF_5	-35.96	-25.37 ± 0.27	-5.99 ± 0.40	-3.24 ± 0.33	0.00 ± 0.45	0.17 ± 0.07	25.50 ± 8.39	2.31 ± 1.71
						0.20 ± 0.03	84.74 ± 18.76	2.80 ± 2.85
						0.16 ± 0.06	300.15 ± 101.78	4.35 ± 1.82
						0.13 ± 0.04	903.27 ± 391.54	3.93 ± 0.58
REF_10	-40.94	-64.01 ± 0.29	-33.49 ± 0.35	-16.00 ± 0.42	0.00 ± 0.37	0.16 ± 0.02	36.70 ± 16.28	3.50 ± 2.85
						0.15 ± 0.03	154.11 ± 61.49	4.34 ± 1.52
						0.16 ± 0.05	338.09 ± 112.05	2.28 ± 1.17
						0.16 ± 0.03	1342.12 ± 207.09	3.87 ± 0.74
THE_5	-81.69	-72.10 ± 0.27	-33.75 ± 0.29	-21.61 ± 0.31	0.00 ± 0.28	0.11 ± 0.01	21.90 ± 0.29	1.54 ± 0.74
						0.15 ± 0.01	100.87 ± 0.34	3.25 ± 0.21
						0.10 ± 0.02	202.06 ± 2.23	5.32 ± 0.28
						0.14 ± 0.01	292.43 ± 2.88	1.78 ± 0.21
THE_10	-163.46	-140.08 ± 0.30	-57.63 ± 0.73	-19.08 ± 0.43	0.00 ± 0.45	0.08 ± 0.01	26.03 ± 0.28	3.87 ± 2.81
						0.15 ± 0.01	100.87 ± 0.10	3.24 ± 0.13
						0.10 ± 0.01	196.96 ± 0.55	4.84 ± 0.19
						0.14 ± 0.01	291.99 ± 0.93	1.68 ± 0.14
Space_5	-246.16	-150.70 ± 0.29	-67.53 ± 0.32	-15.77 ± 0.29	0.00 ± 0.42	0.11 ± 0.05	54.62 ± 38.23	2.61 ± 1.06
						0.14 ± 0.03	140.49 ± 48.03	3.80 ± 0.71
						0.10 ± 0.02	221.09 ± 25.28	3.38 ± 1.95
						0.14 ± 0.02	289.50 ± 3.67	1.59 ± 0.27
Space_10	-478.19	-308.00 ± 0.33	-113.51 ± 0.30	-99.73 ± 0.47	0.00 ± 0.50	0.06 ± 0.01	23.04 ± 0.47	3.54 ± 0.23
						0.16 ± 0.06	100.84 ± 0.06	3.34 ± 0.08
						0.11 ± 0.01	197.35 ± 0.35	4.62 ± 0.12
						0.14 ± 0.01	292.60 ± 0.60	1.74 ± 0.09

Table A.9: As in TableA.6 but for System 4. The results are discussed in section 3.9.4

	$\Delta \ln R_{\text{favoured}} = 0$							
Schedule	$N_p = 0$	$N_p = 1$	$N_p = 2$	$N_p = 3$	$N_p = 4$	RV Semi Amplitude / $\mathrm{m\,s^{-1}}$	Period / days	Phase / radians
True						N/A	N/A	N/A
REF_5	0.00	-3.14 ± 0.32	-3.52 ± 0.22	-5.01 ± 0.21	-6.79 ± 0.20			
REF_10	0.00	-33.33 ± 0.23	-28.32 ± 0.23	-25.11 ± 0.24	-19.73 ± 0.25			
THE_5	-40.08	-10.13 ± 0.24	-7.04 ± 0.24	0.00 ± 0.31	-5.79 ± 0.27	0.07 ± 0.01	16.57 ± 1.14	0.85 ± 0.96
						0.08 ± 0.01	23.98 ± 15.87	4.88 ± 0.58
						0.08 ± 0.01	379.34 ± 39.94	2.10 ± 0.68
THE_10	-19.43	-9.90 ± 0.30	-10.61 ± 0.28	-2.89 ± 0.34	0.00 ± 0.43	0.07 ± 0.01	25.78 ± 1.05	2.77 ± 2.78
						0.07 ± 0.01	30.74 ± 4.30	2.98 ± 2.67
						0.07 ± 0.02	544.98 ± 412.99	2.99 ± 1.68
						0.06 ± 0.02	1007.50 ± 233.98	3.37 ± 0.96
Space_5	-110.53	-19.56 ± 0.26	-10.43 ± 0.35	-3.19 ± 0.39	0.00 ± 0.38	0.06 ± 0.01	137.24 ± 14.58	4.94 ± 0.55
						0.06 ± 0.01	180.91 ± 6.80	3.83 ± 0.41
						0.06 ± 0.01	249.32 ± 8.62	0.80 ± 0.74
						0.06 ± 0.01	384.23 ± 32.34	2.41 ± 0.67
Space_10	-54.50	-16.42 ± 0.32	-10.74 ± 0.30	0.00 ± 0.30	-7.48 ± .72	0.06 ± 0.01	22.11 ± 0.01	2.82 ± 0.27
						0.05 ± 0.01	26.49 ± 0.28	1.29 ± 0.23
						0.05 ± 0.01	871.00 ± 15.61	1.93 ± 0.30

References

Aigrain S., Pont F., Zucker S., 2012, MNRAS, 419, 3147

Bakos G. Á., Lázár J., Papp I., Sári P., Green E. M., 2002, Publications of the Astronomical Society of the Pacific, 114, 974

Baranne A., et al., 1996, A&AS, 119, 373

Beaulieu J.-P., et al., 2006, Nature, 439, 437

Berdyugina S. V., 2005, Living Reviews in Solar Physics, 2, 8

Bolmont E., Libert A.-S., Leconte J., Selsis F., 2016, A&A, 591, A106

Brown T. M., Noyes R. W., Nisenson P., Korzennik S. G., Horner S., 1994, Publications of the Astronomical Society of the Pacific, 106, 1285

Burdanov A., Delrez L., Gillon M., Jehin E., 2018, SPECULOOS Exoplanet Search and Its Prototype on TRAPPIST. Springer International Publishing AG, p. 130, doi:10.1007/978-3-319-55333-7_130

Campbell B., Walker G. A. H., Yang S., 1988, ApJ, 331, 902

Catanzarite J., Shao M., 2011, ApJ, 738, 151

Cersullo F., Coffinet A., Chazelas B., Lovis C., Pepe F., 2019, A&A, 624, A122

Chaplin W. J., Cegla H. M., Watson C. A., Davies G. R., Ball W. H., 2019, AJ, 157, 163

Charbonneau D., Brown T. M., Latham D. W., Mayor M., 2000, ApJ, 529, L45

Coffinet A., Lovis C., Dumusque X., Pepe F., 2019, arXiv e-prints, p. arXiv:1901.03294

Cosentino R., et al., 2012, in Ground-based and Airborne Instrumentation for Astronomy IV. p. 84461V

Crouzier A., 2014, PhD thesis, UJF

Crouzier A., et al., 2012, in Optical and Infrared Interferometry III. p. 84451P (arXiv:1208.0099), doi:10.1117/12.926147

Damon P. E., Jirikowic J. L., 1992, Radiocarbon, 34, 199

Desort M., Lagrange A. M., Galland F., Udry S., Mayor M., 2007, A&A, 473, 983

Dumusque X., 2016, A&A, 593, A5

Dumusque X., Udry S., Lovis C., Santos N. C., Monteiro M. J. P. F. G., 2011a, A&A, 525, A140

Dumusque X., Santos N. C., Udry S., Lovis C., Bonfils X., 2011b, A&A, 527, A82

Dumusque X., Boisse I., Santos N. C., 2014, ApJ, 796, 132

Dumusque X., Boisse I., Santos N. C., 2015a, SOAP 2.0: Spot Oscillation And Planet 2.0, Astrophysics Source Code Library (ascl:1504.021)

Dumusque X., Pepe F., Lovis C., Latham D. W., 2015b, ApJ, 808, 171

Dumusque X., et al., 2015c, ApJ, 814, L21

Dumusque X., et al., 2017, A&A, 598, A133

Feng F., Tuomi M., Jones H. R. A., Barnes J., Anglada-Escudé G., Vogt S. S., Butler R. P., 2017, AJ, 154, 135

Feroz F., Hobson M. P., 2008, MNRAS, 384, 449

Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601

Feroz F., Balan S. T., Hobson M. P., 2011, MNRAS, 415, 3462

Feroz F., Hobson M. P., Cameron E., Pettitt A. N., 2013, arXiv e-prints

Fischer D. A., et al., 2016, PASP, 128, 066001

Foreman-Mackey D., Hogg D. W., Morton T. D., 2014, ApJ, 795, 64

Gaia Collaboration et al., 2016, Astronomy and Astrophysics, 595, A1

Gaia Collaboration et al., 2018, Astronomy and Astrophysics, 616, A1

Gillon M., et al., 2017, Nature, 542, 456

Gregory P. C., 2007a, MNRAS, 374, 1321

Gregory P. C., 2007b, Monthly Notices of the Royal Astronomical Society, 374, 1321

Günther M. N., Queloz D., Demory B.-O., Bouchy F., 2017, MNRAS, 465, 3379

Hall J. C., 2008, Living Reviews in Solar Physics, 5, 2

Hall R., Thompson S., Queloz D., 2016, SPIE Proceedings High Energy, Optical, and Infrared Detectors for Astronomy VII, 9915

Hall R. D., Thompson S. J., Handley W., Queloz D., 2018, MNRAS, 479, 2968

Halverson S., et al., 2016, in Ground-based and Airborne Instrumentation for Astronomy VI. p. 99086P (arXiv:1607.05634), doi:10.1117/12.2232761

Handley W. J., Hobson M. P., Lasenby A. N., 2015a, MNRAS, 450, L61

Handley W. J., Hobson M. P., Lasenby A. N., 2015b, MNRAS, 453, 4384

Hathaway D., 2019, Solar Cycle Science, Available at http://solarcyclescience.com/solarcycle.htmll (2019/03/02)

Haywood R. D., 2015, PhD thesis, University of St Andrews, doi:10.5281/zenodo.35161

Haywood R. D., et al., 2014, MNRAS, 443, 2517

Haywood R. D., et al., 2016, MNRAS, 457, 3637

Herrero E., Ribas I., Jordi C., Morales J. C., Perger M., Rosich A., 2016, A&A, 586, A131

Hilditch R. W., 2001, An Introduction to Close Binary Stars. Cambridge University Press

Howell S. B., et al., 2014, PASP, 126, 398

Huélamo N., et al., 2008, A&A, 489, L9

Janesick J., 2001, Scientific Charge-coupled Devices. Press Monograph Series, Society of Photo Optical

Jeffreys H., 1983, Theory of Probability. International series of monographs on physics, Clarendon Press, https://books.google.co.uk/books?id=EbodAQAAMAAJ

Jurgenson C., Fischer D., McCracken T., Sawyer D., Szymkowiak A., Davis A., Muller G., Santoro F., 2016, in Ground-based and Airborne Instrumentation for Astronomy VI. p. 99086T (arXiv:1606.04413)

Koch D. G., et al., 2010, ApJ, 713, L79

Lagrange A.-M., et al., 2009, A&A, 493, L21

Lagrange A.-M., Meunier N., Desort M., Malbet F., 2011, A&A, 528, L9

Lanza A. F., et al., 2018, A&A, 616, A155

Lanza A. F., Collier Cameron A., Haywood R. D., 2019, arXiv e-prints, p. arXiv:1904.05608

Léger A., et al., 2009, A&A, 506, 287

Li C.-H., et al., 2008, Nature, 452, 610

Lomb N. R., 1976, Ap&SS, 39, 447

Lovis C., et al., 2006a, Nature, 441, 305

Lovis C., et al., 2006b, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. p. 62690P, doi:10.1117/12.669991

MacKay D. J., 2003, Information theory, inference and learning algorithms. Cambridge university press

Mayor M., Queloz D., 1995, Nature, 378, 355

Mayor M., et al., 2003, The Messenger, 114, 20

McMillan R. S., Moore T. L., Perry M. L., Smith P. H., 1993, ApJ, 403, 801

Messenger E., 2012, Astronomical Spectrograph Calibration at the Exo-Earth Detection Limit, Available at http://www.eso.org/sci/publications/messenger/archive/no.149-sep12/messenger-no149-2-6.pdf

Meunier N., Lagrange A.-M., 2013, A&A, 551, A101

Mulders G. D., Pascucci I., Apai D., Ciesla F. J., 2018, AJ, 156, 24

Murphy M. T., et al., 2007, MNRAS, 380, 839

NExSci 2019, NASA Exoplanet Archive, Available at https://exoplanetarchive.ipac.caltech.edu

Noyes R. W., Hartmann L. W., Baliunas S. L., Duncan D. K., Vaughan A. H., 1984, ApJ, 279, 763

Paulson D. B., Yelda S., 2006, PASP, 118, 706

Pepe F., Mayor M., Galland F., Naef D., Queloz D., Santos N., Udry S., Burnet M., 2002, Astronomy & Astrophysics, 388, 632

Pepe F., et al., 2014, Astronomische Nachrichten, 335, 8

Perryman M., 2014, The Exoplanet Handbook. Cambridge University Press

Pesnell W. D., Thompson B. J., Chamberlin P. C., 2012, Sol. Phys., 275, 3

Queloz D., Casse M., Mayor M., 1999, in Hearnshaw J. B., Scarfe C. D., eds, Astronomical Society of the Pacific Conference Series Vol. 185, IAU Colloq. 170: Precise Stellar Radial Velocities. p. 13

Queloz D., et al., 2000, A&A, 354, 99

Queloz D., et al., 2001, A&A, 379, 279

Queloz D., et al., 2009, A&A, 506, 303

Quintana E. V., et al., 2014, Science, 344, 277

Rajpaul V., Aigrain S., Roberts S., 2016, MNRAS, 456, L6

Ravi A., et al., 2017, Journal of Astronomical Telescopes, Instruments, and Systems, 3, 045003

Reiners A., Banyal R. K., Ulbrich R. G., 2014, A&A, 569, A77

Ricker G. R., et al., 2014, in Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave. p. 914320

Robertson P., Mahadevan S., Endl M., Roy A., 2014, Science, 345, 440

Rossum G., 1995, Technical report, Python Tutorial. Centre for Mathematics and Computer Science, Amsterdam, The Netherlands, The Netherlands

Scargle J. D., 1982, ApJ, 263, 835

Seager S., Kuchner M., Hier-Majumder C. A., Militzer B., 2007, The Astrophysical Journal, 669, 1279

Seager S., Bains W., Petkowski J. J., 2016, Astrobiology, 16, 465

Shaklan S., Sharman M. C., Pravdo S. H., 1995, Appl. Opt., 34, 6672

Shapiro A. I., Solanki S. K., Krivova N. A., Schmutz W. K., Ball W. T., Knaack R., Rozanov E. V., Unruh Y. C., 2014, A&A, 569, A38

Sing D. K., et al., 2016, Nature, 529, 59

Skilling J., 2006, Bayesian Anal., 1, 833

Solanki S. K., 2003, The Astronomy and Astrophysics Review, 11, 153

Street R. A., et al., 2003, in Deming D., Seager S., eds, Astronomical Society of the Pacific Conference Series Vol. 294, Scientific Frontiers in Research on Extrasolar Planets. pp 405–408 (arXiv:astro-ph/0208233)

Struve O., 1952, The Observatory, 72, 199

Thompson S. J., et al., 2016, in Ground-based and Airborne Instrumentation for Astronomy VI. p. 99086F

Tompsett M. F., Amelio G. F., Smith G. E., 1970, Applied Physics Letters, 17, 111

Tuomi M., et al., 2013, A&A, 551, A79

VanderPlas J. T., 2018, The Astrophysical Journal Supplement Series, 236, 16

Voelz D. G., 2011, Computational Fourier Optics: A MATLAB Tutorial. SPIE, doi:10.1117/3.858456, https://doi.org/10.1117/3.858456

Vogt S. S., 1992, in European Southern Observatory Conference and Workshop Proceedings. p. 223

Vogt S. S., et al., 2014, PASP, 126, 359

Wakeford H. R., et al., 2018, AJ, 155, 29

Wang Y., Liu Y., Tian F., Hu Y., Huang Y., 2017, preprint (arXiv:1710.01405)

Wheatley P. J., et al., 2018, MNRAS, 475, 4476

Wildi F., Pepe F., Chazelas B., Curto G. L., Lovis C., 2011, in Techniques and Instrumentation for Detection of Exoplanets V. p. 81511F

Winn J. N., 2010, arXiv e-prints, p. arXiv:1001.2010

Winn J. N., Fabrycky D. C., 2015, Annual Review of Astronomy and Astrophysics, 53, 409

Zeng L., et al., 2018, arXiv e-prints, p. arXiv:1806.05909